AMC PAMPHLET AMCP 706-198

AMCP 706-198 DEVELOPMENT GUIDE FOR RELIABILITY, PART FOUR JANUARY 1976

ENGINEERING DESIGN HANDBOOK

DEVELOPMENT GUIDE FOR RELIABILITY

PART FOUR

RELIABILITY MEASUREMENT

HEADQUARTERS, US ARMY MATERIEL COMMAND

DEPARTMENT OF THE ARMY
HEADQUARTERS UNITED STATES ARMY MATERIEL COMMAND
5001 Eisenhower Ave., Alexandria, VA 22333

AMC PAMPHLET
NO. 706-198

7 January 1976

DEVELOPMENT GUIDE FOR RELIABILITY
PART FOUR
RELIABILITY MEASUREMENT

TABLE OF CONTENTS

TABLE OF CONTENTS (Con't.)

CHAPTER 3. STATISTICAL EVALUATION OF RELIABILITY TESTS. DEMONSTRATIONS. AND ACCEPTANCE

TABLE OF CONTENTS (Con't.)

TABLE OF CONTENTS (Con't.)

TABLE OF CONTENTS (Con't.)

TABLE OF CONTENTS (Con't.)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Con't.)

LIST OF TABLES

LIST OF TABLES (Con't.)

LIST OF TABLES (Con't.)

PREFACE

This handbook, *Reliability Measurement,* is the third in a series of five on reliability. The series is directed largely toward the working engineers who have the responsibility for creating and producing equipment and systems which can be relied upon by the users in the field.

The five handbooks are:

1. *Design for Reliability,* AMCP 706-196
2. *Reliability Prediction,* AMCP 706-197
3. *Reliability Measurement,* AMCP 706-198
4. *Contracting for Reliability,* AMCP 706-199
5. *Mathematical Appendix and Glossary,* AMCP 706-200.

This handbook is directed toward reliability engineers who need to be familiar with statistical analysis of experimental results, with reliability-test management and planning, and with integrated reliability-data systems. Many examples are used, especially to illustrate the statistical analysis. References are given to the literature for further information.

The majority of the handbook content was obtained from many individuals, reports, journals, books, and other literature. It is impractical here to acknowledge the assistance of everyone who made a contribution.

The original volume was prepared by Tracor Jitco, Inc. The revision was prepared by Dr. Ralph A. Evans of Evans Associates, Durham, NC, for the Engineering Handbook Office of the Research Triangle Institute, prime contractor to the US Army Materiel Command. Technical guidance and coordination on the original draft were provided by a committee under the direction of Mr. O. P. Bruno, Army Materiel System Analysis Agency, US Army Materiel Command.

The Engineering Design Handbooks fall into two basic categories, those approved for release and sale, and those classified for security reasons. The US Army Materiel Command policy is to release these Engineering Design Handbooks in accordance with current DOD Directive 7230.7, dated 18 September 1973. All unclassified Handbooks can be obtained from the National Technical Information Service (NTIS). Procedures for acquiring these Handbooks follow:

a. All Department of Army activities having need for the Handbooks must submit their request on an official requisition form (DA Form 17, dated Jan 70) directly to:

Commander
Letterkenny Army Depot
ATTN: AMXLE-ATD
Chambersburg, PA 17201

(Requests for classified documents must be submitted, with appropriate "Need to Know" justification, to Letterkenny Army Depot.) DA activities will not requisition Handbooks for further free distribution.

b. All other requestors, DOD, Navy, Air Force, Marine Corps, nonmilitary Government agencies, contractors, private industry, individuals, universities, and others must purchase these Handbooks from:

National Technical Information Service
Department of Commerce
Springfield, VA 22151

Classified documents may be released on a "Need to Know" basis verified by an official Department of Army representative and processed from Defense Documentation Center (DDC), ATTN: DDC-TSR, Cameron Station, Alexandria, VA 22314.

Comments and suggestions on this Handbook are welcome and should be addressed to:

Commander
US Army Materiel Development and Readiness Command
Alexandria, VA 22333

(DA Forms 2028, Recommended Changes to Publications, which are available through normal publications supply channels, may be used for comments/suggestions.)

CHAPTER 1

INTRODUCTION

1-1 GENERAL

Reliability measurement techniques provide a common discipline that can be used to make system reliability projections throughout the life cycle of a system. The data on component and equipment failures obtained during the reliability measurement program can be used to compute component failure distributions and equipment reliability characteristics. Reliability measurement techniques are used during the research and development phase to measure the reliability of components and equipments and to evaluate the relationships between applied stresses and environments and reliability. Later in a system life cycle, reliability measurement and testing procedures can be used to demonstrate that contractually required reliability levels have been met.

Uniform criteria for establishing a reliability measurement program are defined in MIL-STD-785 (Ref. 1). These standards must be incorporated into Department of Defense procurements of all systems that undergo contract definition. If a system does not require a contract definition effort, they can be incorporated in the request for proposal (RFP).

The Army has developed a number of regulations, complementing MIL-STD-785, which establish reliability as a major parameter which must be measured during the development of a new weapon system (Refs. 2, 3, and 4). All Army materiel ought to be physically tested to determine whether the design requirements, including reliability, have been met. Testing is performed under the direction of the appropriate AMC commodity commands, project managers, and installations or activities which report directly to Headquarters AMC.

The US Army Test and Evaluation Command (USATECOM) is responsible for reviewing test documentation produced by other Army organizations. USATECOM can, at its own discretion, conduct independent tests and evaluations on any Army developed system (Ref. 5). The reliability measurement techniques described in this volume are consistent with Army Regulations and can be applied directly to systems developed under Army auspices.

A reliability measurement system consists of two major functional divisions: (1) the test program, and (2) the data system.

The test program provides a comprehensive test effort that ensures that reliability goals are met. A test schedule that designates when test procedures, test samples, and necessary equipment and facilities will be required must be developed. Procedures for gathering the data, which will be generated throughout all phases of the test program, must be documented in sufficient detail for complete identification and integration into the data processing system.

The integrated data system establishes procedures for accumulating, coding, and handling data. Standard data accumulation requirements (and compatible data sheets) provide for collecting and recording data such as identifying-information, environmental conditions, operating hours or cycles, and failures.

1-2 RELIABILITY TEST PROGRAM

Testing is an important and expensive part of the development program for an equip ment or system. Because reliability testing is expensive, the test program must be designed carefully to fit into the overall development program.

The information provided by a reliability test program can be used at any point in the system life cycle. A well planned program of functional environmental testing must be conducted during system design and development in order to achieve the required reliability and to provide data for improving reliability. These tests are used to measure the reliability of components, equipments, and subsystems used in a system. ,Tests also are used to design tools to evaluate the relationship between various environments and stresses, and reliability. The reliability tests performed during development answer the basic question of whether or not the design really works.

A reliability measurement program must be planned carefully. The contractor developing a system must prepare an integrated test plan that includes all reliability tests to be performed during the program. The tests must be designed to make maximum use of all data produced on the program. The reliability test program must be integrated with other system/equipment test programs in order to minimize wasted effort. A number of standard test plans have been developed to guide contractors and Army project managers (Refs. *6*, 7, and 8). The test plans in MIL-STD-781 can be used for testing equipments and systems whose failure characteristics are governed by the exponential distribution. The sampling plans described in MIL-STD-105 can be used for 1-shot devices. Modifications of the procedures in MIL-STD-105 for components governed by a Weibull distribution are described in TR-7 (Ref. 4).

After the design is established, reliability tests can be used to make decisions about system reliability and to determine if reliability goals have been met. The procedures described in this volume can be applied to a variety of situations. The tests range from quality-assurance tests, which are performed at the part level on lots of components, to reliability demonstration tests used to prove that a system indeed meets its reliability requirements.

Demonstration tests on systems can be performed in three distinct phases:

1. Specific subsystems and equipments must be tested to determine if they meet the reliability requirements allocated to them,, The equipments must be evaluated in a controlled environment in which performance is monitored by means of an instrumented test set-up. Equipments that do not meet reliability requirements must be redesigned.

2. After individual equipments have been tested, they must be mated and the entire system must be subjected to realistic operational procedures and environments. Reliability data are gathered by means of a carefully organized data reporting system.

3. Operational testing must be performed by Army personnel who exercise the system in the operational environment. Reliability data are gathered along with many other data items. These tests permit the reliability performance of Army systems to be determined in realistic operating environments, and they may uncover weaknesses masked in the previous tests.

Achieved reliability must be demonstrated formally at specified times during the program. Demonstration testing must be performed at the system level, and at the subsystem and equipment levels. Demonstration test plans must include a definition of failure criteria, applied environments and stresses, test procedures, and the applicable statistical test plans.

The techniques of mathematical statistics are used extensively in reliability testing. These techniques provide the tools that relate sample size, test duration, s-confidence levels, stress levels, and other factors. They are discussed in Chapters 2 and 3, and in *Part Six, Mathematical Appendix and Glossary.*

Chapter 4 will describe techniques and procedures which can be applied to reliability test planning and management to ensure a more efficient test program.

1-3 INTEGRATED DATA SYSTEM

An integrated data system can be used to provide project managers and engineers with the data that they need in order to monitor the reliability achieved by the system and its component parts. If provided in a timely manner, this information can be used for effective planning, review, and control of actions related to system reliability. The data system ought to encompass the following characteristics:

1. Be a closed-loop system for collecting, analyzing, and recording all failures that occur during system development.

2. Provide data that can be used to estimate reliability and to indicate needed corrective action. All hardware failures should be recorded with information about the failed component, time of failure, cause of failure, and other pertinent information.

3. Develop computer programs that permit the printing of reliability status output reports.

4. Develop and standardize procedures for data accumulation and reduction. These standard procedures must provide for the collection of data, the recording of identifying information, environmental conditions, operating hours or cycles, and hardware failures on each test performed.

5. Be structured to make use of data recorded on failures that occur at times other than the reliability tests.

6. Handle, process, and integrate all data obtained from testing, inspection, and failure trouble reporting. These data can be used for reliability analysis and reporting, assessment of equipment readiness, and a variety of other purposes.

7. Maintain and update a computer data bank of accumulated reliability data. These data can be processed to produce reliability status reports that present a summary of failure rates and reliability parameters for components, equipments, and subsystems. These reports can be structured to list the troublesome items that are causing the most serious reliability difficulties. They can be distributed to cognizant Army and contractor engineers and managers.

A detailed description of an integrated data system is presented in Chapter 5. Chapters 6, 7, and 8 describe environmental testing, accelerated testing, and nondestructive testing, respectively.

REFERENCES

1. MIL-STD-785, *Reliability Program for Systems and Equipment Development and Production.*

2. AMCR 11-1, *Research and Development, Systems Analysis.*

3. AR 702-3, *Product Assurance: Army Materiel Reliability, Availability, and Maintainability (RAM).*

4. TR-7, *Factors and Procedures for Applying MIL-STD-105D Sampling Plans to*

Life and Reliability Testing, Office of the Assistant Secretary of Defense for Installations and Logistics. Available from NTIS.

5. USATECOM Pamphlet 700-700, *USATECOM Materiel Test Procedures.*

6. MIL-STD-781, *Reliability Tests: Exponential Distribution.*

7. MIL-STD-105, *Sampling Procedures and Tables for Inspection by Attributes.*

CHAPTER 2

STATISTICAL EVALUATION OF RELIABILITY TESTS, DESIGN AND DEVELOPMENT

LIST OF SYMBOLS

B = function of β in lognormal distribution, see par. 2-3.6

c = acceptance number (for sampling); number of cycles

C_i = s-confidence limit

Cdf = Cumulative distribution function, $Cdf\{x\} \equiv Pr\{X \leqslant x\}$

Conf$\{\cdot\}$ = s-confidence that the statement in the $\{\cdot\}$ is true

Cov$\{\cdot\}$ = Covariance

$csqf(\chi^2; \nu)$ = Cdf of chi-square distribution with ν degrees of freedom

$csqfc(\chi^2; \nu)$ = complement of $csqf(\chi^2; \nu)$, $csqf(\chi^2; \nu) = 1 - csqf(\chi^2; \nu)$

$E\{\cdot\}$ = sexpected value, mean

F = a Cdf; the F-statistic

$gauf(z)$ = Cdf of Gaussian (s-normal) distribution

$gaufc(z)$ = complement of $gauf(z)$, $gaufc(z) = 1 - gauf(z)$

H = cumulative hazard, $Sf = \exp(-H)$

i = an integer in Example No. 10 (cell boundary number)

k = number of events

k_i = number of trials with result i

K-S = critical value in a K-S test (see Table 2-13)

L, U = subscripts meaning Lower and Upper

m, b = slope and intercept for y, see par. 2-8

n_0, n_e = see **Eqs.** 2-43 and 2-44

N = sample size

o = subscript, implies a fixed value—not a random variable

OC = Operating Characteristic

p_i = probability of result i (constant from trial to trial)

$p_{k,N}$ = k-th from N order-statistic

pdf = probability density function

$poif(i; \mu)$ = Cdf of the Poisson distribution (mean is μ)

$poim\ (i;\mu)$ = *pmf* of the Poisson distribution (mean is μ)

pmf = probability mass function

PP = plotting position

r = number of failures

R = an *Sf*

s- implies the word "statistical(ly)", or implies that the technical statistical definition is intended rather than the ordinary dictionary definition

s^2 = SS/ν, variance estimate

S_i = i/N, basic plotting position, see par. 2-5

Sf = Survivor function, $Sf\{x\} \equiv Pr\{X \geqslant x\}$

SS = sum of squares

t = time; time-to-failure; Student t variable

T = total test time

$\mathrm{Var}\{\cdot\}$ = Variance, square of standard deviation

$weif\ (u;\beta)$ = *Cdf* of the standard Weibull distribution (shape parameter is β)

x = random variable

$\bar{x}$ = mean of a sample of x's

X^2 = see Eqs. 2-43 and **2-44**

Y = a linear function of **x**, see par. 2-8

α = scale parameter; producer risk

β = shape parameter; consumer risk

Γ = gamma function

δ = a difference associated with a sample, see par. 2-9

ϵ = standard s-normal variate

η = coefficient of variation, $\eta \equiv \sigma/\mu$

θ = exponential scale parameter; mean time to failure

λ = Poisson rate parameter; failure rate

μ = mean; also location-parameter of s-normal distribution

ν = degrees of freedom

ρ = linear correlation coefficient

σ = standard deviation; also scale-parameter of s-normal distribution

ϕ = fraction of mission time, see Eq. 2-68

χ^2 = a random variable which has the chi-square distribution

χ^2_ν = same as χ^2, but implies the associated degrees of freedom

$\chi^2_{p,\nu}$ = same as χ^2_ν, but also implies the probability (*Cdf*) associated with that value of χ^2_ν, viz., $csqf\ (\chi^2_{p,\nu};\nu) = p$.

implies the value before a shift of the origin, see par. 2-8

implies an estimated value of the parameter

2-1 INTRODUCTION

The main advantage of statistics is that it can provide a good measure of the uncertainty involved in a numerical analysis. The secondary advantage is that it does provide methods for estimating effects that might otherwise be lost in the random variations in the data. Wherever possible in the examples in this chapter, the uncertainty in the numerical results will be emphasized.

Rarely is an engineer interested only in the results of analyzing his model. The engineer must solve a real-world, not a mathematical, problem. The answer to the mathematical problem must be tempered by all the other important considerations that never found their way into the model; this is why the estimation of uncertainty is so important. The engineer needs to know how much he can afford to be swayed by those other considerations.

A well designed and properly executed reliability test program provides useful data for system designers and managers. The statistical tests described in this chapter can be used to help ensure that the system design meets reliability requirements. A description of the basic concepts of statistical testing during system design and development is presented in this chapter.

Reliability test and measurement are among the most important parts of a design and development program (Ref. 1). During design and development, tests are performed to:

1. Measure the reliability of equipments and subsystems (measurement tests)

2. Evaluate the relationships between applied environments and stresses and reliability (evaluation tests)

3. Verify that an item meets a prescribed minimum reliability (tests of verification)

4. Select the more reliable unit or approach from several alternatives (tests of comparison).

Reliability-measurement tests must be conducted under controlled conditions that approximate those to which the equipment will be subjected in the field (Ref. 2). Operating times and number of failures are accumulated and used to estimate the underlying failure distribution, the reliability, and level of s-confidence of the results.

Evaluation tests provide estimates of the relationships between failures, and applied environments and stresses. Numerical relationships between failure rate (and reliability) and specific stresses can be derived. In addition, the relative effects of each environment in a multienvironment situation can be estimated using techniques such as Analysis of Variance and Multiple Regression.

Tests of verification are used to verify that a desired result has been obtained (Ref. 2). A hypothesis such as "the reliability is equal to or greater than 0.95 for 500 hr of operation" or "the failure rate is equal to or less than 0.02 per 1000 hr" is tested. The test hypothesis is then verified at some level of s-significance by the test results. A wide variety of tests can be designed—depending on the number of units tested, the time allowed for testing, and the level of risks taken in accepting the results.

Frequently, alternate design approaches are available to the system designer. Tests of comparison permit the designer to compare their reliability. The basic hypothesis of this kind of test is that "no difference exists between the reliabilities". This hypothesis is tested against the hypothesis that "the reliabilities are not equal". These tests provide useful guidance to equipment designers who can then make decisions based on test results and rigorous statistical analyses.

The tests must be well planned and test data properly evaluated in order to avoid costly errors and delays. This is especially true for system reliability testing in which components frequently are destroyed and in which expensive equipment must be built to simulate the operational environment. Test planning is very important in complex programs that operate under time and budget limitations. Critical trade-offs must be made among test time, number of units tested, and achieved s-confidence level.

Reliability measurement tests are used to make estimates of the reliability of a population of items. Both parametric and nonparametric estimates can be used. Parametric estimates are based on a known or assumed distribution of the characteristic of interest. The constants in the equation that describe the probability distribution are called parameters. Nonparametric estimates are made without assuming any particular form for the probability distribution.

The three types of parametric estimates most frequently used are (Ref. 3):

1. Point estimate—a single-valued estimate of a reliability parameter

2. Interval estimate—an estimate of an interval that is believed to contain the true value of the parameter

3. Distribution estimate—an estimate of the parameters of a reliability distribution.

A s-confidence interval estimate is one for which there is a known probability that the true value of the unknown parameter or characteristic lies within a computed interval. s-Confidence interval estimates are more useful than point estimates because they give a much better idea of the uncertainty involved in the estimation process.

Distribution estimates are used when it is desired to estimate the probability distribution governing a particular reliability measure. This is usually a 2-step process: (1) the form of the distribution must be hypothesized or determined from the failure data, and (2) the parameters that describe the distribution must be estimated.

Nonparametric methods can be used to estimate reliability measures without making any assumptions concerning the time-to-failure distribution. Generally, nonparametric estimates are not as efficient as parametric estimates. Nonparametric reliability estimates apply only to the specific test interval and cannot be extrapolated. Both point estimates and interval estimates can be made using nonparametric techniques.

Before one decides to choose a type of distribution and estimate its parameters from the data, one should have a very good idea of why it is being done. For example, one may wish to use it to interpolate among the data, one may wish to extrapolate, or one may wish to estimate certain characteristics of the data such as a mean, median, or tenth percentile. If one is going to use the distribution to interpolate among the data, goodness-of-fit tests are quite appropriate to help determine a good type of distribution to use. If one is going to extrapolate, then goodness-of-fit tests (which operate only in the region of the data) are not appropriate because they do not tell how well the distribution will fit in the region where there are no data; in fact, goodness-of-fit tests for this purpose can be extremely misleading.

If the purpose of knowing the distribution is to estimate some characteristics of the population, one should give serious consideration to calculating the corresponding sample property and using that directly to estimate the population property. It is essentially a distribution-free method and is not subject to errors caused by the distribution not fitting the population out in the tail region where there are no data. Goodness-of-fit tests for this purpose are appropriate if the only property of the distribution that is being used is one inside the data. It is quite inappropriate if properties of the distribution outside the data are used, for example, in calculating a mean.

2-2 GRAPHICAL ESTIMATION OF PARAMETERS OF A DISTRIBUTION

The underlying distribution governing the reliability characteristics should be chosen carefully, because the validity of the reliability predictions and tests depends on this selection. Although the exponential distribution is most common for electronic equipments, other distributions are used. The failure characteristics of electromechanical and mechanical systems often can be described by distributions such as the s-normal, lognormal, or Weibull. Computer programs are available for estimating the parameters for an assumed distribution from a set of data. In many practical cases, graphical techniques are simple to apply and produce adequate results for estimating the underlying distribution. They are virtually always a useful preliminary to analytic estimation.

The basic idea in developing special graph paper for use in graphical analysis is to have the population *Cdf* or its cumulative hazard plot as a straight line, A straight line has 2 parameters (slope and intercept); so 2 parameters for the distribution can be determined, if the distribution can be appropriately transformed.

Graphical curve fitting techniques have been developed for all of the distributions commonly associated with reliability testing (Refs. 3, 4, 5, 22, and 23). Procedures for the s-normal, lognormal, and Weibull distributions are quite simple to apply, and are illustrated in the remainder of this paragraph.

In graphical methods, the data from the sample are rearranged so that they are in order from smallest to largest; they are then referred to as order-statistics. Occasionally the order is from largest to smallest, but since it is so rarely done in reliability work, all illustrations will be in the usual way. In order to plot the data points, a method is needed for choosing the *Cdf* (or the equivalent cumulative hazard) at which each point is to be plotted. There are several methods for doing this; they are explained in par. 2-2.1 and 2-2.2. There is no clear-cut way that is acceptable to everyone. But some of the disagreements are needless, for the simple reason that when the sample size is small, the inherent uncertainty in plotting position is very large (regardless of the method used); and when the sample size is large, all the methods tend to give the same position. Besides, if the finer phases of parameter estimation by graphical methods are important to you, you ought to be using an analytic method for those finer phases—graphical methods just don't have the ability to make precise point estimates or to estimate the uncertainty in those point estimates.

2-2.1 PLOTTING POSITIONS (CUMULATIVE DISTRIBUTION FUNCTION)

Two methods of determining plotting position are described. Both require that once testing is stopped for any nonfailed item, it be stopped for all remaining items. Likewise, neither can use the extra information if testing continues beyond the failure time of the last recorded item (this tends to be true for any graphical method and many analytic methods).

1. The sample *Cdf* is plotted, and the uncertainty is looked-up in a simple table.

If all items are failed, the probability statements about the true *Cdf* (relative to the sample points) are simple and straightforward. This method is recommended whenever it can be used.

2. The distribution of the order statistics is used to provide **3** plotting positions **for** each point. This spread provides a feel for the uncertainties involved. Simple, straightforward probability statements can be made only for an individual point, not for the *Cdf* as a whole. Extensive tables are necessary.

As mentioned in the first paragraph, these methods can be used provided no failure time is longer than a censoring time. **A** censoring (censored item) occurs when an item is removed from test before it fails. The cause for removal cannot be related to the apparent condition of the item if an analysis in Chapter 2 is to be valid. For example, suppose as failure becomes quite likely, an item begins to vibrate slightly. Then if items that vibrate are removed from test before they actually fail, the removal cause is related to the condition of the item; and a legitimate analysis of test results is virtually impossible unless the whole test and population description are redefined.

2-2.1.1 Practical Plotting (K-S Bounds)

Specific instructions for this kind of plotting are on the back of each sheet of Practical Probability Paper in par. 2-2. They are repeated here for a general case.

Notation:

F = Cumulative distribution function (*Cdf*)
n = sample size
r = failure number; $r = 1,2, \ldots, n$

Plotting instructions follow:

1. Plotting data: plot failure r at the two points

$$F_{Hi} = r/n \tag{2-1a}$$

$$F_{Lo} = (r-1)/n \tag{2-1b}$$

Connect the points with horizontal and vertical lines; this is the sample *Cdf*.

2. 2-sided s-confidence bounds on the actual *Cdf*: choose the s-confidence level, near $[1-(1/n)]$ is reasonable; then find **KS,** from the body of Table 2-1(A) (e.g., $n = 10$, s-confidence = 95%, **KS,** = 0.41). The upper bound is plotted at

$$F_{Lo} + \text{KS}, \text{ and } F_{Hi} + \text{KS}_n \ ; \tag{2-2a}$$

the lower bound is plotted at

$$F_{Hi} - \text{KS}, \text{ and } F_{Lo} - \text{KS}, \ . \tag{2-2b}$$

For each bound, connect the points with

TABLE 2-1(A)

TABLE OF K-S BOUNDS

n	KS, (s-confidence) (90%)	(95%)	(98%)	(99%)
5	.51	.56	.63	.67
6	.47	.52	.58	.62
8	.41	.45	.51	.54
10	.37	.41	.46	.49
12	.34	.38	.42	.45
14	.31	.35	.39	.42
16	.30	.33	.37	.39
18	.28	.31	.35	.37
20	.26	.29	.33	.36
30	.22	.24	.27	.29
40	.19	.21	.24	.25
n	$\frac{1.22}{\sqrt{n+1}}$	$\frac{1.36}{\sqrt{n+1}}$	$\frac{1.52}{\sqrt{n+1}}$	$\frac{1.63}{\sqrt{n+1}}$

(formula is o.k. for $n \geqslant 6$)

For censored samples, use an n which is between the original-number-on-test and the number-of-failures.

horizontal and vertical lines; they will be parallel to, and **KS,** from, the sample *Cdf*. Then "1 – s-confidence" is the fraction of times you-do-this-procedure that the true *Cdf* will partly lie outside the 2-sided *s*-confidence bounds. In general, you will be disheartened at how little you know about the true *Cdf*.

Drawing the data-lines: draw the two parallel lines, farthest apart, that fit reasonably well within the s-confidence bounds; use both to estimate bounds on the "intercept" parameter of the straight line (e.g., the mean for the s-normal distribution). Draw the two intersecting lines, with steepest and smallest slopes, that fit reasonably well within the s-confidence bounds; use both to estimate bounds on the "slope" parameter of the straight line (e.g., the standard deviation for a s-normal distribution).

Table 2-1(A) also can be used the other way: if a true *Cdf* is drawn, then all sample points will lie within $\pm KS_n$ from it, with the stated s-confidence. Several examples are given in the paragraphs that follow for using this method of **K-S** Bounds. The **K-S** stands for Kolmogorov-Smirnov (two Russians who developed much of the theory).

2-2.1.2 Plotting (Beta Bounds)

If a sample of N is drawn from the uniform distribution (representing a *Cdf*), and then the results are put in order from lowest to highest, the *pdf* and Cdf of the k-th order-statistic are

$$Cdf\{p_{k,N}\} = \sum_{j=k}^{N} \binom{N}{j} p_{k,N}^{j} (1 - p_{k,N})^{N-j} \tag{2-3a}$$

$$pdf\{p_{k,N}\} = N\binom{N-1}{k-1} p_{k,N}^{k-1} (1 - p_{k,N})^{N-k} \tag{2-3b}$$

This is a beta distribution and its properties are well known (see ***Part Six, Mathematical Appendix and*** *Glossary*). The mean, mode, and median of $p_{k,N}$ are

$$E\{p_{k,N}\} = k/(N+1) \tag{2-4a}$$

$$\text{mode}\{p_{k,N}\} = (k-1)/(N-1) \tag{2-4b}$$

$$\text{median}\{p_{k,N}\} \cong (k-0.3)/(N+0.4) \tag{2-4c}$$

The mean value, **Eq.** 2-4a, is used often and reasonably as the plotting position–simply because it is so easy to calculate; but it, alone, gives no idea of the uncertainty involved due to the random nature of the data. The median, **Eq.** 2-4c is also reasonable to use, but is less popular because of its greater complexity. The expression $(k - 0.5)/N$ is fairly popular as a plotting position, and is as reasonable as any single one, but it has no simple property to give it a name.

Table 2-1(B) lists 3 plotting positions for each point; so the uncertainty is plainly shown on the graph. These are the points for which the *Cdf* in **Eq.** 2-3a is 5%, 50%, 95%; i.e., only 1 point in 10 would be outside that range. Later paragraphs in this chapter illustrate the use of these plotting positions.

This method of plotting is called "Beta Bounds", just to have a short name for it.

2-2.2 PLOTTING POSITIONS (CUMULATIVE HAZARD)

When some of the items are removed from test before they fail, and the test is continued for other failures, the plotting positions for the Cdf are very difficult to calculate. Ref. 23 shows how data can be simply plotted in this situation. The errors incurred in using this method are probably small compared to the uncertainties involved. Unfortunately, it is not feasible to provide a rigorous measure of that uncertainty.

TABLE 2-1(B)

BETA BOUNDS METHOD: PERCENTAGE PLOTTING POINTS OF THE k-th ORDERED-FAILURE, OUT OF A TOTAL SAMPLE OF *N*.

k \ N	3	4	5	6	7	8	9	10	11
1	1.7/21/63	1.2/16/53	1.0/13/45	0.85/11/39	0.74/9.4/35	0.65/8.3/31	0.57/7.4/28	0.51/6.7/26	0.47/6.1/24
2	14/50/86	9.8/39/75	7.6/31/66	6.3/27/58	5.3/23/52	4.7/20/47	4.1/18/43	3.7/16/39	3.3/15/36
3			19/50/81	15/42/73	13/36/66	11/32/60	9.8/29/55	8.7/26/51	8.0/24/47
4					23/50/77	19/44/71	17/39/66	15/36/61	14/32/56
5							25/50/75	22/45/70	20/41/65
6									27/50/73

k \ N	12	13	14	15	16	17	18	19	20
1	0.43/5.6/22	0.40/5.2/21	0.37/4.8/19	0.34/4.5/18	0.32/4.2/17	0.30/4.0/16	0.29/3.8/16	0.28/3.6/15	0.26/3.4/14
2	3.1/14/34	2.8/13/32	2.6/12/30	2.5/11/28	2.3/10/26	2.2/9.8/25	2.1/9.2/24	1.9/8.7/23	1.8/8.3/22
3	7.2/22/44	6.7/20/41	6.1/19/39	5.7/18/36	5.4/16/34	5.0/16/33	4.8/15/31	4.5/14/30	4.3/13/29
4	12/30/53	11/28/49	10/26/47	9.7/24/44	9.1/23/42	85/21/40	8.0/20/38	7.6/19/36	7.3/18/35
5	18/38/61	17/35/57	15/33/54	14/31/51	13/29/48	12/27/46	12/26/44	11/24/42	11/23/40
6	25/46/68	23/43/65	21/40/61	19/37/58	18/35/55	17/33/52	16/31/50	15/29/48	14/28/46
7		29/50/71	27/47/67	25/44/64	23/41/61	21/39/58	20/36/55	19/35/53	18/33/51
8				30/50/70	28/47/67	26/44/65	24/42/61	23/40/58	22/38/56
9						31/50/69	29/47/66	27/45/63	26/43/60
10								32/50/68	30/48/65

k \ N	22	24	26	28	30	35	40	45	50
1	0.23/3.1/13	0.21/2.9/12	0.20/2.6/11	0.18/2.5/10	0.17/2.3/9.5	0.15/2.0/8.2	0.13/1.7/7.2	0.11/1.5/6.4	0.10/1.4/5.8
2	1.6/7.5/20	1.5/6.9/18	1.4/6.4/17	1.3/5.9/16	1.2/5.5/15	1.0/4.8/13	0.90/4.2/11	0.79/3.7/10	0.71/3.3/9.1
3	3.8/12/26	3.5/11/24	3.2/10/22	3.0/9.4/21	2.8/8.8/20	2.4/7.6/17	2.1/6.6/15	1.8/5.9/13	1.7/5.3/12
4	6.5/16/32	5.9/15/29	5.4/14/27	5.0/13/25	4.7/12/24	4.0/10/21	3.5/9.1/18	3.1/8.1/16	2.8/7.3/15
5	9.4/21/37	8.6/19/34	7.9/18/32	7.3/16/30	6.8/15/28	5.8/13/24	5.1/12/21	4.5/10/19	4.0/9.3/17
6	13/25/42	11/23/39	11/22/36	9.8/20/34	9.1/19/32	7.7/16/28	6.7/14/25	6.0/13/22	5.4/11/20
7	16/30/47	15/27/43	13/25/40	12/24/38	12/22/36	9.8/19/31	8.5/17/27	7.5/15/25	6.8/13/22
8	20/34/52	18/32/48	16/29/45	15/27/42	14/25/39	12/22/34	10/19/30	9.2/17/27	8.3/15/25
9	23/39/56	21/36/52	19/33/49	18/31/46	17/29/43	14/25/37	12/21/33	11/19/30	9.7/17/27
10	27/43/60	25/40/56	23/37/53	21/34/49	19/32/47	16/27/41	14/24/36	13/21/32	11/19/29
11	31/48/65	28/44/60	26/41/56	24/38/53	22/35/50	19/30/44	16/26/39	14/24/35	13/21/32
12		32/48/64	29/44/60	27/41/57	25/38/53	22/34/47	18/29/41	16/26/37	14/23/34

TABLE 2-1(B). (cont'd)

	22	24	26	28	30	35	40	45	50
13			33/48/64	30/45/60	28/42/57	24/36/50	20/31/44	18/28/40	16/25/36
14				33/48/63	31/45/60	26/39/52	22/34/47	20/30/42	18/27/38
15					34/48/63	29/42/55	25/36/49	22/32/44	19/29/40
16						31/44/58	27/39/52	24/35/47	21/31/42
17						34/47/61	29/41/54	26/37/49	23/33/44
18						36/50/64	31/44/57	28/39/51	25/35/47
19							34/46/59	30/41/53	26/37/49
20							36/49/62	32/43/56	28/39/51
21								34/46/58	30/41/53
22								36/48/60	32/43/55
23								38/50/62	34/45/57
24									36/47/59
25									38/49/60

The body of the table lists for each (k, N) the 5%/50%/95% points for plotting purposes. To obtain the 5%/50%/95% plotting points for $(N + 1 - k, N)$ reverse the order from the (k, N) and subtract each from **100%**. For example, for $(k, N) = (2, 5)$ the percentage plotting points are **7.6/31/66**. For $(N + 1 - k, N) = (4,5)$, the percentage plotting points are **(100 – 66)/(100 – 31)/(100 – 7.6)** = 34/69/92.4.

Points through n = 20 are adapted from Ref. 4

Points above n = 20 are adapted from Ref. **26.**

All are rounded off to 2 significant figures.

Interpolation for values of **N** not shown: For k small, interpolate (roughly) on a horizontal line. For values of k near $(N/2)$, interpolate on a diagonal ($k/N \approx$ constant); in that region they are roughly of the form: median plotting-point ± deviation, The deviation is easily calculated from the tabulated values, and the median plotting-point is easily estimated from Eq. 2-5.

The cumulative hazard H is related to the Sf by the equations

$$Sf\{x\} = \exp\ [-H(x)] \tag{2-5a}$$

$$H(x) = -\ln\ [Sf\{x\}] = -\ln\ [1 - Cdf\{x\}] \tag{2-5b}$$

So a plotting scale can be calculated for any probability paper by using Eq. 2-5b. Special paper can be drawn for more convenient hazard plotting and is mentioned in Ref. 23; but it is not at all necessary.

Even if no rigorous method is available for estimating the uncertainty, it is desirable to get some idea about it. The procedure that follows provides grossly-approximate **K-S** bounds (par. 2-2.1.1). It has the advantage that the same general plotting technique is used as for the **K-S** Bounds method. These instructions are also given on the Instructions side of the Practical Probability Paper.

Plot failure r at the two points

$$F_{Hi} = 1 - \exp\ (-H_r) \tag{2-6a}$$

$$F_{Lo} = 1 - \exp\ (-H_{r-1}) \tag{2-6b}$$

to convert the sample cumulative hazard H_r to the Cdf. Connect the points with horizontal and vertical lines; this is the sample Cdf. Calculate and plot the **K-S** bounds as in par. 2-2.1.1 and **Eq.** 2-2. The s-confidence bounds will not be exact at all.

Table 2-2 shows some failure data on field windings of electric generators. The hazard

TABLE 2-2

CUMULATIVE-HAZARD CALCULATIONS FOR FIELD WINDINGS OF SOME ELECTRIC GENERATORS

The table lists time-to-failure (months) for failed units and time-to-end-of-test (months) for unfailed units. All times are listed in order of increasing time.

Rank r	Ordered event time	Reverse rank	Hazard increment ΔH	Cumulative hazard H_r
1	31.7 F	16	0.0625	0.0625
2	39.2 F	15	0.0667	0.1292
3	57.5 F	14	0.0714	0.201
4	65.0	13		
5	65.8 F	12	0.0833	0.284
6	70.0 F	11	0.0909	0.375
7	75.0	10		
8	75.0	9		
9	87.5	8		
10	88.3	7		
11	94.2	6		
12	101.7	5		
13	105.8 F	4	0.250	0.625
14	109.2	3		
15	110.0 F	2	0.500	1.12
16	130.0	1		

F indicates failure. Other times are censorings (removed from test before failure, for a reason not connected with the state of the item).

increment is the reciprocal of the reverse rank for the failed units; the hazard does not increment for nonfailed units. The times-to-failure are plotted on the time scale of the appropriate paper, and the cumulative hazard is plotted on the cumulative-hazard or on the *Cdf* scale. This procedure is illustrated in some of the examples.

2-2.3 *s*-NORMAL DISTRIBUTION

The basis for the graph paper is that the equation

$$F \equiv Cdf\{x\} = \text{gauf}\left(\frac{x-\mu}{\sigma}\right) \tag{2-7}$$

can be transformed to

$$gauf^{-1}(F) = \frac{x-\mu}{\sigma} \tag{2-8a}$$

or

$$x = \sigma \cdot gauf^{-1}(F) + \mu \ . \tag{2-8b}$$

The *Cdf* scale is actually the $gauf^{-1}$ scale. The heavy dashed line on *s*-Normal Practical Probability Paper is drawn as

$$F = 50\%, \ x = \mu \ . \tag{2-9a}$$

The S-scale is linear in $gauf^{-1}$. For example, the S = ±1 points are at

$$F = 84.13\%, \ x = \mu + \sigma, \ \text{for } S = +1 \tag{2-9b}$$

$$F = 15.87\%, \ x = \mu - \sigma, \ \text{for } S = -1. \tag{2-9c}$$

Several examples illustrating methods of plotting the data are presented—Example Nos. 1(A) through 1(F). The data are prepared 3 ways:

1. For K-S Bounds plotting
2. For Beta Bounds plotting
3. To show the Hazard plotting technique.

Since the data are not censored (they are complete), the Hazard plotting is not necessary. But the comparison of plotting positions is useful. Col. 8 (plotting position for Hazard plotting) is very close to col. 3 (plotting position for K-S Bounds method). Cols. 5-8 merely illustrate the calculations for the Hazard plotting and are not used elsewhere.

First the K-S Bounds Method is shown. s-Normal Practical Probability Paper is used. Cols. 2-3 from Table 2-3 (Data Set A) are used; see Example No. 1(A).

It is not feasible to put quantitative s-confidence levels on the interval estimates for μ and σ; analytic methods are necessary for that. Nevertheless, these intervals are a good engineering measure of the uncertainties involved. Some of the important conclusions from this graphical exercise are:

1. Not very much is known about the distribution of failure times of these fuel pumps. For example, if one were interested in the time at which 2% of the fuel pumps will have failed, it is tempting to use line #5, and guess about 240 hr. But that point really is only known to within the range 0 to 1200 hr.

2. Whether the data can reasonably be represented (summarized) by an s-Normal distribution is almost irrelevant.

3. 5% of the time we go through this procedure (the s-confidence was 95%), the true *Cdf* will not lie wholly within that very wide envelope.

4. The fuel pump may have roughly 10% defectives (line #4), i.e., lives so short as to be of critical concern. Perhaps this is reason enough to ground the aircraft until further

TABLE 2-3

DATA SET A (Ref. 4)

Ordered failure times (in flight hours) for a pump in the fuel-delivery system of an aerial fire-support helicopter. N = 20. The entire set of 20 was failed; so there is no censoring. Columns 1 and 2 present the original data. Column 3 shows the calculation for plotting the sample Cdf in the K-S Bounds method. Column 4 shows the 3 plotting positions for the Beta Bounds method. The remaining columns are included merely to show how the Hazard plotting compares with *Cdf* plotting in the case where there is no censoring and the two methods can be compared directly.

Rank	Failure time, hr	r/N, %	Cdf plotting position, % (Table 2-1(B))	Reverse rank	Hazard increment $\Delta H^{(1,2)}$	Cumulative hazard $H_r^{(1,3)}$	$Cdf = 1 - \exp(-H_r)$, %$^{(4)}$
1	175	5	0.26/3.4/14	20	0.050	0.050	4.9
2	695	10	1.8/8.3/22	19	0.053	0.103	9.8
3	872	15	4.3/13/29	18	0.056	0.158	14.6
4	1250	20	7.3/18/35	17	0.059	0.217	19.5
5	1291	25	11/23/40	16	0.063	0.280	24.4
6	1402	30	14/28/46	15	0.067	0.346	29.3
7	1404	35	18/33/51	14	0.071	0.418	34.1
8	1713	40	22/38/56	13	0.077	0.495	39.0
9	1741	45	26/43/60	12	0.083	0.578	43.9
10	1893	50	30/48/65	11	0.091	0.669	48.8
11	2025	55	35/52/70	10	0.100	0.769	53.6
12	2115	60	40/57/74	9	0.111	0.880	58.5
13	2172	65	44/62/78	8	0.125	1.005	63.4
14	2418	70	49/67/82	7	0.143	1.148	68.3
15	2583	75	54/72/86	6	0.167	1.314	73.1
16	2725	80	60/77/89	5	0.200	1.514	78.0
17	2844	85	65/82/92.7	4	0.250	1.764	82.9
18	2890	90	71/87/95.7	3	0.333	2.098	87.7
19	3268	95	78/91.7/98.2	2	0.500	2.598	92.6
20	3538	100	86/96.6/99.74	1	1.000	3.598	97.3

1) Calculations made to 8 significant figures; only 3 decimal places recorded.

2) AH = 1/(reverse rank)

3) $H_r = \Sigma \Delta H_i$

4) This is the *Cdf* plotting position that corresponds to the cumulative hazard; this technique is used when the Hazard scale is not shown on the graph paper.

investigation and/or corrective action shows it to be safe. The long-lived units might be studied to find out why they were so good.

5. If the K-S Bounds method had not been used, the uncertainty would not have been realized. *An* engineer could easily have presumed that line #5 was the whole story and thus misled himself and others about the results.

The same data set will now be plotted by the Beta Bounds method. **Cols.** 2 and **4** from Table **2-3** (Data Set **A)** are used; see Example No. 1 (B).

Two more data sets (B in Table 2-4, and C in Table 2-5) are plotted to help illustrate

Example No. 1(A)

Data Set A, K–S Bounds Method (Table 2–3, Fig. 2–1(A))

Procedure	Example
1. Choose a s-confidence level. Use a number near $1 - 1/N$. Find KS_n from Table 2–1(A).	1. $N = 20$ $1 - 1/N = 95\%$. Use 95% s-confidence. $KS_n = 0.29$.
2. Plot the data from Cols. 2–3 using the instructions on the Practical Probability Paper.	2. Prepare Col. 3 of Table 2-3. Plot on Fig. 2–1(A).
3. Find the lower and upper estimates of μ, μ_L and μ_U.	3. Lines #1 and #2 are the two parallel lines for μ. #1 intersects the heavy dashed line (50% line) at $\mu_L = 1400$ hr, #2 at $\mu_U = 2440$ hr.
4. Find the lower and upper estimates of σ, σ_L and σ_U.	4. Lines #3 and #4 are the two intersecting lines for σ. For line #3, choose $S_2 = +1$, $S_1 = 0$; then $\sigma_L = 2250$ hr $-$ 1960 hr = 290 hr. For line #4, choose $S_2 = +0.5$, $S_1 = -0.5$; then $\sigma_U = 3360$ hr $-$ 400 hr = 2960 hr.
5. Draw line #5 for the point estimates of μ and σ, $\hat{\mu}$ and $\hat{\sigma}$.	5. Lines #3 and #4 were drawn so that their intersection would be at the 50% line and midway between lines #1 and #2. This was for an unnecessary esthetic sense, so that line #5 could be drawn through the intersection of lines #3 and #4 and be parallel to and midway between lines #1 and #2. $\hat{\mu} = 1920$ hr, $\hat{\sigma} = (2700 \text{ hr} - 1080 \text{ hr})/[+1 - (-1)]$ = 810 hr.
6. Summarize the results.	6. $\mu_U = 2440$ hr, $\sigma_U = 2960$ hr $\hat{\mu} = 1920$ hr, $\hat{\sigma} = 810$ hr $\mu_L = 1400$ hr, $\hat{\sigma}_L = 290$ hr.

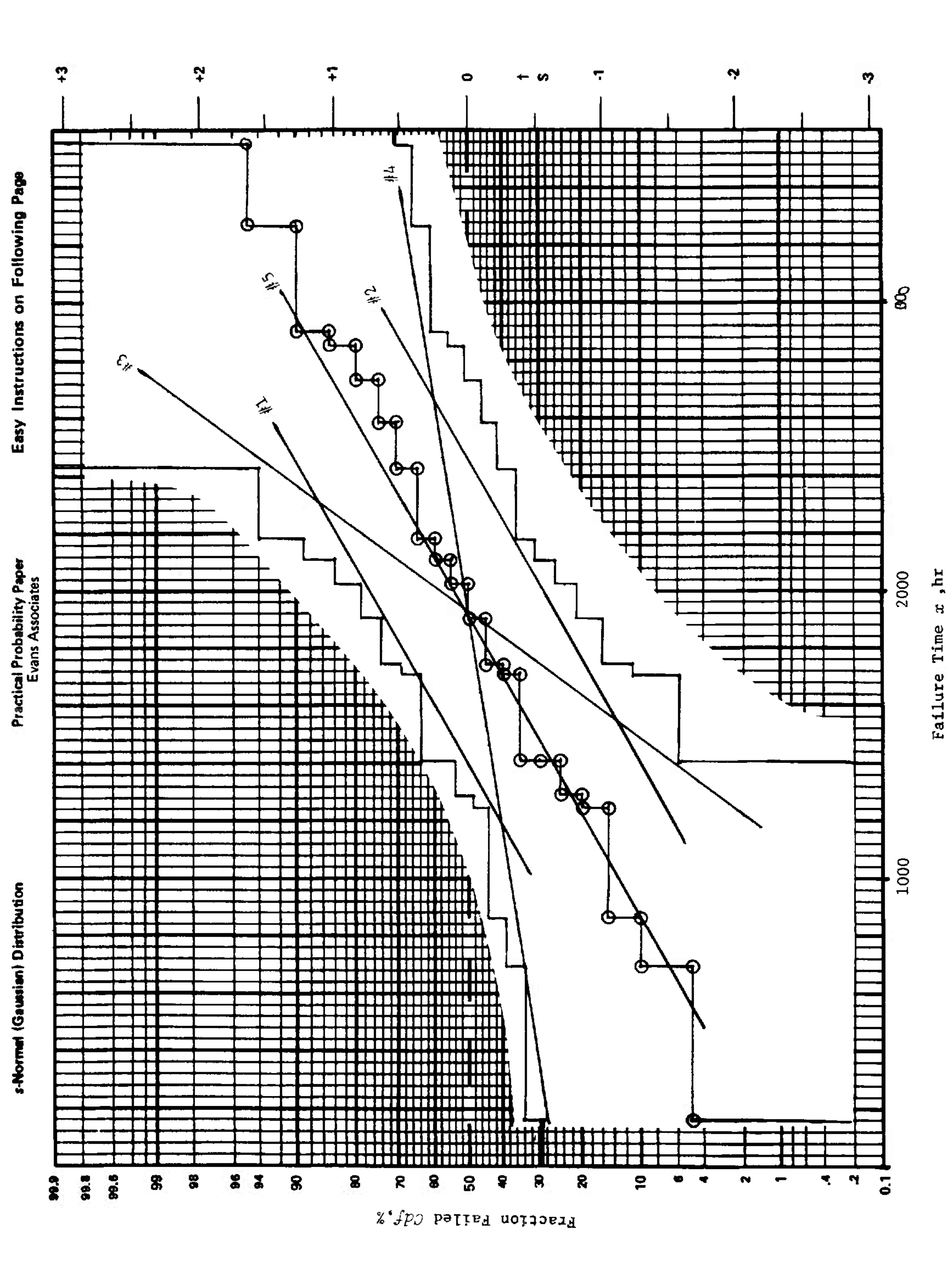


Figure 2-1 (A). Data Set A, s-Normal Scale. K-S Bounds Method

Practical Probability Paper
Evans Associates

Instructions for Use

s-Normal (Gaussian) Distribution

- Plotting data: Plot failure r at the two points: $F_{Hi} = r/n$ and $F_{Lo} = (r-1)/n$. Connect the points with horizontal and vertical lines; this is the sample Cdf.

- Plotting cumulative-hazard data: Plot failure r at the two points $F_{Hi} = 1 - \exp(-H_r)$ and $F_{Lo} = 1 - \exp(-H_{r-1})$ to convert the sample cumulative-hazard H_r to the Cdf. Connect the points with horizontal and vertical lines; this is the sample Cdf. The s-confidence bounds will not be exact.

- 2-sided sconfidence bounds on the actual Cdf: Choose the s-confidence level, near $1-(1/n)$ is reasonable; then find KS_n from the body of the Table (e.g., n=10, s-conf=95%, KS_n=0.41). The upper bound is plotted at $F_{Lo}+KS_n$ and $F_{Hi}+KS_n$; the lower bound is plotted at $F_{Hi}-KS_n$ and $F_{Lo}-KS_n$. For each bound, connect the points with horizontal and vertical lines; they will be parallel to, and KS_n from, the sample Cdf. Then 1—s-conf is the fraction of times youdo-this-procedure that the true Cdf will partly lie outside the 2-sided s-confidence bounds. In general, you will be disheartened at how little you know about the true Cdf.

- Drawing data-lines: Draw the two parallel lines, farthest apart, that fit reasonably within the s-confidence bounds; use both to estimate bounds on μ. Draw the two intersecting lines, with steepest and smallest slopes, that fit reasonably within the s-confidence bounds; use both to estimate bounds on σ.

- To estimate μ: μ is the value of x at which a data-line intersects the heavy dashed line (F_μ = 50.0%).

- To estimate σ: Take two values of S, S_1 and S_2; then find the two x values, x_1 and x_2, which correspond to S_1 and S_2, via a data-line. $\sigma = |x_2 - x_1|/|S_2 - S_1|$. If $|S_2 - S_1| = 1$, then $\sigma = |x_2 - x_1|$.

Table of K-S Bounds

n	KS_n (sconfidence)			
	(90%)	(95%)	(98%)	(99%)
5	.51	.56	.63	.67
6	.47	.52	.58	.62
8	.41	.45	.51	.54
10	.37	.41	.46	.49
12	.34	.38	.42	.45
14	.31	.35	.39	.42
16	.30	.33	.37	.39
18	.28	.31	.35	.37
20	.26	.29	.33	.36
30	.22	.24	.27	.29
40	.19	.21	.24	.25
n	$\frac{1.22}{\sqrt{n+1}}$	$\frac{1.36}{\sqrt{n+1}}$	$\frac{1.52}{\sqrt{n+1}}$	$\frac{1.63}{\sqrt{n+1}}$

(formula is o.k. for $n \geqslant 6$)

For censored samples, use an n which is between the original-number-on-test and the number-of-failures.

s-Normal (Gaussian) Distribution

$F(x) = \text{gauf}\left(\frac{x-\mu}{\sigma}\right)$, gauf (z) is the standard s-normal Cdf

F Cumulative distribution function (Cdf)

μ location parameter (same units as x), also the median and mean (average)

σ scale parameter (same units as x), also the standard deviation

n sample size

r failure number; r = 1, 2, ... ,n

Example No. 1(B)

Data Set A, Beta Bounds Method (Table 2–3, Fig. 2-1(B))

Procedure	Example
1. Plot the 5%/50%/95% points for each failure time.	1. Prepare Table 2–3 col. 3 from Table 2–1(B). Plot on Fig. 2–1(B).
2. Sketch a curve through the 5% and through the 95% plotting points.	2. *See* curves 1 and 5 on Fig. 2–1(B).
3. Draw the "best" straight line through the 50% plotting points. It will more or less bisect the region between the 5% and 95% plotting point curves.	3. Draw line #3.
4. Draw 2 more straight lines through a central 50% population point: one with the maximum feasible slope and the other with the minimum feasible slope.	4. Draw lines #2 and #4.
5. Estimate μ_U, $\hat{\mu}$, μ_L from the curves #1, #3, #5, respectively. They are all the intersection with the heavy dashed line. They are the high limit, point estimate, and low limit, respectively.	5. μ_U = 2460 hr (high limit) $\hat{\mu}$ = 1950 hr (point estimate) μ_L = 1460 hr (low limit).
6. A separate scale is rarely given for the scale parameter σ. It is easily estimated from the fact that the 16% and 84% population points are each 1 σ away from the 50% population point. The S-scale could have been used, along with the instructions on the back of Fig. 2–1(A). (σ_U, $\hat{\sigma}$, and σ_L are read from curves 2, 3, and 4 of Fig. 2–1(B), respectively.)	6. $\sigma_U = (3400 - 500)/2 = 1450$ hr $\hat{\sigma} = (2980 - 1000)/2 = 990$ hr $\sigma_L = (2600 - 1300)/2 = 650$ hr.

It is not feasible to put quantitative s-confidence levels on the upper and lower limits for μ or σ. The conclusions to be drawn are essentially the same as from Fig. 2–1(A).

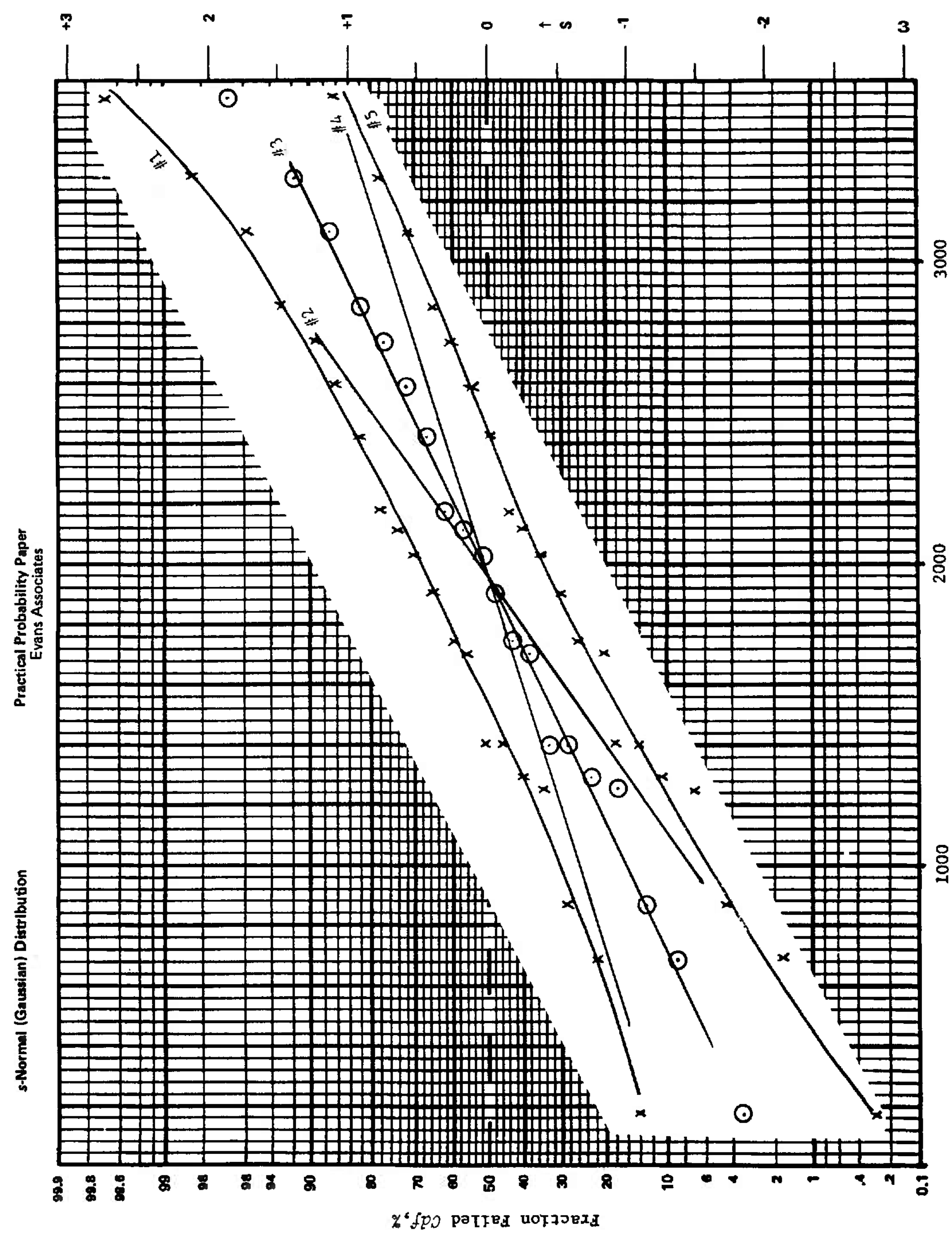


Figure 2-1 (B). Data Set A, s-Normal Scale. Beta Bounds Method

TABLE 2-4

DATA SET B

Ordered failure times (simulated) of ball bearings. Test stopped at the 5th failure, 5 unfailed. $N = 10$.

Rank	Failure time, hr	r/N, %	*Cdf* Plotting position, % (Table 2-1)
1	497	10	0.51/6.7/26
2	546	20	3.7/16/39
3	557	30	8.7/26/51
4	673	40	15/36/61
5	789	50	22/45/70

TABLE 2-5

DATA SET C

Ordered failure times (simulated) of ball bearings. Entire set of 10 was failed. (This is the same set as B, except the test was continued until all had failed.) $N = 10$.

Rank	Failure time, hr	r/N, %	*Cdf* Plotting position, % (Table 2-1)
1	497	10	0.51/6.7/26
2	546	20	3.7/16/39
3	557	30	8.7/26/51
4	673	40	15/36/61
5	789	50	22/45/70
6	805	60	30/55/78
7	1150	70	39/64/85
8	1450	80	49/74/91.3
9	1690	90	61/84/96.3
10	3090	100	74/93.3/99.49

$1 - 1/N = 90\%$; use 90% s-confidence. $KS_n = 0.37$

the utility of the graph paper. Data Set B is censored as shown in Table 2-4; the 4 columns have been prepared as for Data Set **A**, and the points plotted; see Example Nos. 1(C) and 1(D).

The same data set is plotted by the Beta Bounds method for comparison with the **K-S** Bounds method; see Example No. 1(E).

2-2.4 WEIBULL DISTRIBUTION

The form of the Weibull distribution being used is

$$R \equiv Sf\{t; \alpha, \beta\} = \exp\left[-(t/\alpha)^{\beta}\right] \qquad (2\text{-}10)$$

Eq. 2-5b shows that the cumulative hazard for the Weibull is

$$H(t) = (t/\alpha)^{\beta} \qquad (2\text{-}11)$$

The Weibull paper is derived by taking log (ln) of Eq. 2-10.

$$\ln R = -(t/\alpha)^{\beta} \qquad (2\text{-}12a)$$

$$\log(-\ln R) = \beta \log t - \beta \log \alpha \qquad (2\text{-}12b)$$

So log t plotted against log (−ln R) is a straight line. Most Weibull papers use ln $[-\ln(R)]$ for **Eq.** 2-12b and so have to use awkward methods to find β.

The procedure for trying the Weibull distribution is quite similar to that for the s-normal, except that Weibull probability paper is used. This is shown in detail for Data Set D later in this paragraph.

Data Set **A** from Table 2-3 is plotted in Fig. 2-4; the **K-S** Bounds method is used, with 95% s-confidence bounds ($N = 20$, **KS,** $= 0.29$). Following the instructions on the reverse side of Fig. 2-4, one obtains

$$\alpha_L = 1700 \text{ hr}, \ \hat{\alpha} = 2200 \text{ hr},$$

$$\alpha_U = 2900 \text{ hr}$$

$$\beta_L = 0.9, \ \hat{\beta} = 2.1, \ \beta_U = 6 \ .$$

This sample could easily have come from a Weibull distribution. If the true distribution

[text continues on page 2-27]

Example No. 1(C)

This example uses Data Set B, see Example Nos. 1(C) and 1(D), Table 2–4

Data Set B, K–S Bounds Method (Table 2–4, Fig. 2–2)

Procedure	Example
1. Choose a s-confidence level. Use a number near $1 - 1/N$. Find KS_ϵ from Table 2–1(A) or the table on reverse side of Fig. 2–2.	1. $N = 10$ $1 - 1/N = 90\%$. Use 90% s-confidence. $KS_\epsilon = 0.7$.
2. Plot the data from Cols. 2-3 using the instructions on the Practical Probability Paper.	2. Prepare Col. 3 of Table 2–4. Plot on Fig. 2–2.
3. Find the lower and upper estimates of μ, μ_L and μ_U.	3. Use lines #1 and #2. μ $\mu_L = 600$ hr $\mu_U = 1000$ hr.
4. Find the lower and upper estimates of σ, σ_L and σ_U. Use the S scale.	4. Use lines #3 and #4. $\sigma_L = 100$ hr $\sigma_U = 800$ hr.
5. Draw line #5 for the point estimates of μ and σ, $\hat{\mu}$ and $\hat{\sigma}$.	5. $\hat{\mu} = 800$ hr $\hat{\sigma} = 260$ hr.

It is difficult to tell much from the data. The scatter and uncertainty are brought forcibly to the engineer's attention. The 10% failure point appears to be somewhere between 0 and 650 hr. Since the data tell so little, they are not plotted by the Beta Bounds method.

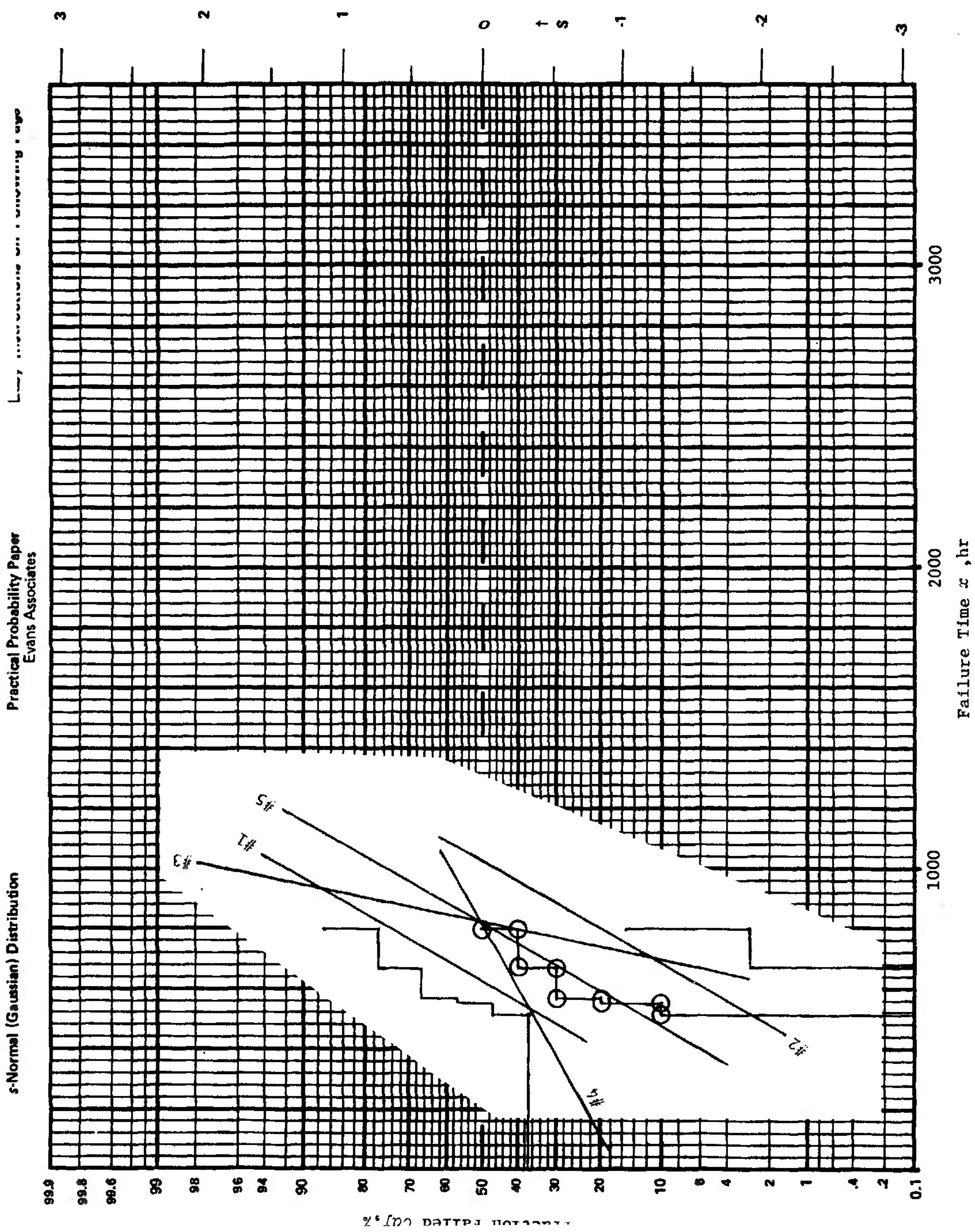


Figure 2-2. Data Set B, s-Normal Scale. K-S Bounds Method

Instructions for Use

s-Normal (Gaussian) Distribution

- Plotting data: Plot failure r at the two points: $F_{Hi} = r/n$ and $F_{Lo} = (r-1)/n$. Connect the points with horizontal and vertical lines; this is the sample Cdf.

- Plotting cumulative-hazard data: Plot failure r at the two points $F_{Hi} = 1 - \exp(-H_r)$ and $F_{Lo} = 1 - \exp(-H_{r-1})$ to convert the sample cumulative-hazard H_r to the Cdf. Connect the points with horizontal and vertical lines; this is the sample Cdf. The s-confidence bounds will not be exact.

- 2-sided s-confidence bounds on the actual Cdf: Choose the s-confidence level, near $1-(1/n)$ is reasonable; then find KS_n from the body of the Table (e.g., n=10, s-conf=95%, KS_n=0.41). The upper bound is plotted at $F_{Lo}+KS_n$ and $F_{Hi}+KS_n$; the lower bound is plotted at $F_{Hi}-KS_n$ and $F_{Lo}-KS_n$. For each bound, connect the points with horizontal and vertical lines; they will be parallel to, and KS_n from, the sample Cdf. Then $1-s$-conf is the fraction of times you-do-this-procedure that the true Cdf will partly lie outside the 2-sided s-confidence bounds. In general, you will be disheartened at how little you know about the true Cdf.

- Drawing data-lines: Draw the two parallel lines, farthest apart, that fit reasonably within the s-confidence bounds; use both to estimate bounds on p. Draw the two intersecting lines, with steepest and smallest slopes, that fit reasonably within the s-confidence bounds; use both to estimate bounds on σ.

- To estimate μ: μ is the value of x at which a data-line intersects the heavy dashed line (F_μ = 50.0%).

- To estimate a: Take two values of S, S_1 and S_2; then find the two x values, x_1 and x_2, which correspond to S_1 and S_2, via a data-line. $a = |x_2 - x_1|/|S_2 - S_1|$. If $|S_2 - S_1| = 1$, then $a = |x_2 - x_1|$.

Table of K-S Bounds

n	KS_n (sconfidence) (90%)	(95%)	(98%)	(99%)
5	.51	.56	.63	.67
6	.47	.52	.58	.62
8	.41	.45	.51	.54
10	.37	.41	.46	.49
12	.34	.38	.42	.45
14	.31	.35	.39	.42
16	.30	.33	.37	.39
18	.28	.31	.35	.37
20	.26	.29	.33	.36
30	.22	.24	.27	.29
40	.19	.21	.24	.25
n	$\frac{1.22}{\sqrt{n+1}}$	$\frac{1.36}{\sqrt{n+1}}$	$\frac{1.52}{\sqrt{n+1}}$	$\frac{1.63}{\sqrt{n+1}}$

(formula is o.k. for $n \geqslant 6$)

For censored samples, use an n which is between the original-number-on-test and the number-of-failures.

s-Normal (Gaussian) **Distribution**

$F(x)$ = gauf $(\frac{x-\mu}{\sigma})$, gauf (z) is the standard s-normal Cdf

F Cumulative distribution function (Cdf)

μ location parameter (same units as x), also the median and mean (average)

σ scale parameter (same units as x), also the standard deviation

n sample size

r failure number; r = 1, 2, ... ,n

Example No. 1(D)

This example uses Data Set C, Table 2–5.

Data Set C, K–S Bounds Method (Table 2–5, Fig. 2–3(A))

Procedure	Example
1. Choose a s-confidence level. Use a number near $1 - 1/N$. Find KS_n from Table 2–1(A) or the table on reverse side of Fig. 2–3(A).	1. N = 10. $1 - 1/N = 90\%$. Use 90% s-confidence. $KS_n = 0.37$.
2. Plot the data from Cols. 2–3 using the instructions on the Practical Probability Paper.	2. Prepare Col. 3 of Table 2–5. Plot on Fig. 2–3(A).
3. Find the lower and upper estimates of μ, μ_L and μ_U.	3. μ_L = 620 hr μ_U = 1400 hr.
4. Find the lower and upper estimates of σ, σ_L and σ_U. Use the S-scale.	4. σ_L = 220 hr σ_U = 1900 hr.
5. Find the point estimates of μ and and σ, $\hat{\mu}$ and $\hat{\sigma}$.	5. $\hat{\mu}$ = 1000 hr σ = 730 hr.

Some of the important conclusions from this graphical exercise are:

1. Not much is known about the shape of the distribution; it could easily be s–Normal.

2. The time for 10% failures is not known well; it is probably between 0 and 800 hr.

3. Only ballpark ideas about the distribution are known.

4. The estimates of the distribution are quite different than from the censored version of Data Set B.

Example No. 1(E)

This example uses Data Set C, Table 2–5.

Data Set *C*, Beta Bounds Method (Table 2–5, Fig. 2–3(B))

Procedure	Example
1. Plot the 5%/50%/95% plotting points.	1. Prepare Col. 4 from Table 2–1(B). Plot the points.
2. Sketch the envelope through the 5% plotting points and through the 95% plotting points.	2. See curves #1 and #3.

There seems little point in going on with this analysis, the lines are so curved that it doesn't seem that the data came from a s-normal distribution. This conclusion is contrary to the ones drawn from the *K–S* Bounds method for the same data. In general, one should tend to believe the *K–S* Bounds method.

In subsequent paragraphs, the Weibull and lognormal distributions will be tried for Data Sets A, B, C to see if they fit better.

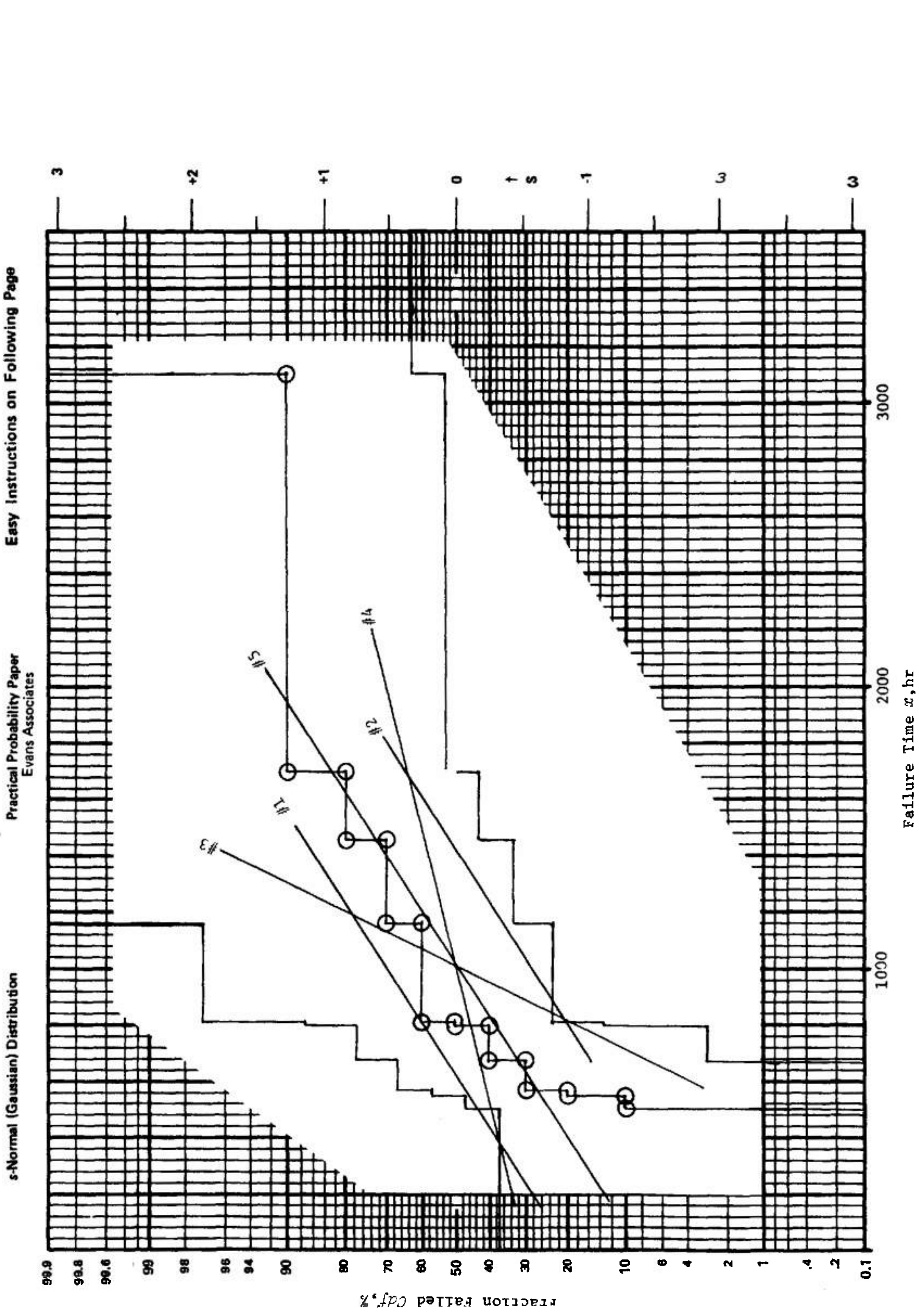


Figure 2-3 (A). Data Set C, s-Normal Scale. K-S Bounds Method

Instructions for Use

s-Normal (Gaussian) Distribution

- Plotting data: Plot failure r at the two points: $F_{Hi} = r/n$ and $F_{Lo} = (r-1)/n$. Connect the points with horizontal and vertical lines; this is the sample Cdf.

- Plotting cumulative-hazard data: Plot failure r at the two points $F_{Hi} = 1-\exp(-H_r)$ and $F_{Lo} = 1-\exp(-H_{r-1})$ to convert the sample cumulative-hazard H_r to the Cdf. Connect the points with horizontal and vertical lines; this is the sample Cdf. The s-confidence bounds will not be exact.

- 2-sided s-confidence bounds 'on the actual Cdf: Choose the s-confidence level, near $1-(1/n)$ is reasonable; then find KS_n from the body of the Table (e.g., n=10, s-conf=95%, KS_n=0.41). The upper bound is plotted at $F_{Lo}+KS_n$ and $F_{Hi}+KS_n$; the lower bound is plotted at $F_{Hi}-KS_n$ and $F_{Lo}-KS_n$. For each bound, connect the points with horizontal and vertical lines; they will be parallel to, and KS_n from, the sample Cdf. Then $1-$s-conf is the fraction of times youdo-this-procedure that the true Cdf will partly lie outside the 2-sided s-confidence bounds. In general, you will be disheartened at how little you know about the true Cdf.

- Drawing data-lines: Draw the two parallel lines, farthest apart, that fit reasonably within the s-confidence bounds; use both to estimate bounds on μ. Draw the two intersecting lines, with steepest and smallest slopes, that fit reasonably within the s-confidence bounds; use both to estimate bounds on σ.

- To estimate μ: μ is the value of x at which a data-line intersects the heavy dashed line (F_μ = 50.0%).

- To estimate σ: Take two values of S, S_1 and S_2; then find the two x values, x_1 and x_2, which correspond to S_1 and S_2, via a data-line. $\sigma = |x_2 - x_1|/|S_2 - S_1|$. If $|S_2 - S_1| = 1$, then $\sigma = |x_2 - x_1|$.

Table of K-S Bounds

n	KS_n (sconfidence) (90%)	(95%)	(98%)	(99%)
5	.51	.56	.63	.67
6	.47	.52	.58	.62
8	.41	.45	.51	.54
10	.37	.41	.46	.49
12	.34	.38	.42	.45
14	.31	.35	.39	.42
16	.30	.33	.37	.39
18	.28	.31	.35	.37
20	.26	.29	.33	.36
30	.22	.24	.27	.29
40	.19	.21	.24	.25
n	$\frac{1.22}{\sqrt{n+1}}$	$\frac{1.36}{\sqrt{n+1}}$	$\frac{1.52}{\sqrt{n+1}}$	$\frac{1.63}{\sqrt{n+1}}$

(formula is o.k. for $n \geqslant 6$)

For censored samples, use an n which is between the original-number+,-test and the number-of-failures.

s-Normal (Gaussian) Distribution

$F(x) = \text{gauf}\left(\frac{x-\mu}{\sigma}\right)$, gauf (z) is the standard s-normal Cdf

F Cumulative distribution function (Cdf)

μ location parameter (same units as x), also the median and mean (average)

σ scale parameter (same units as x), also the standard deviation

n sample size

r failure number; r = 1, 2, ... ,n

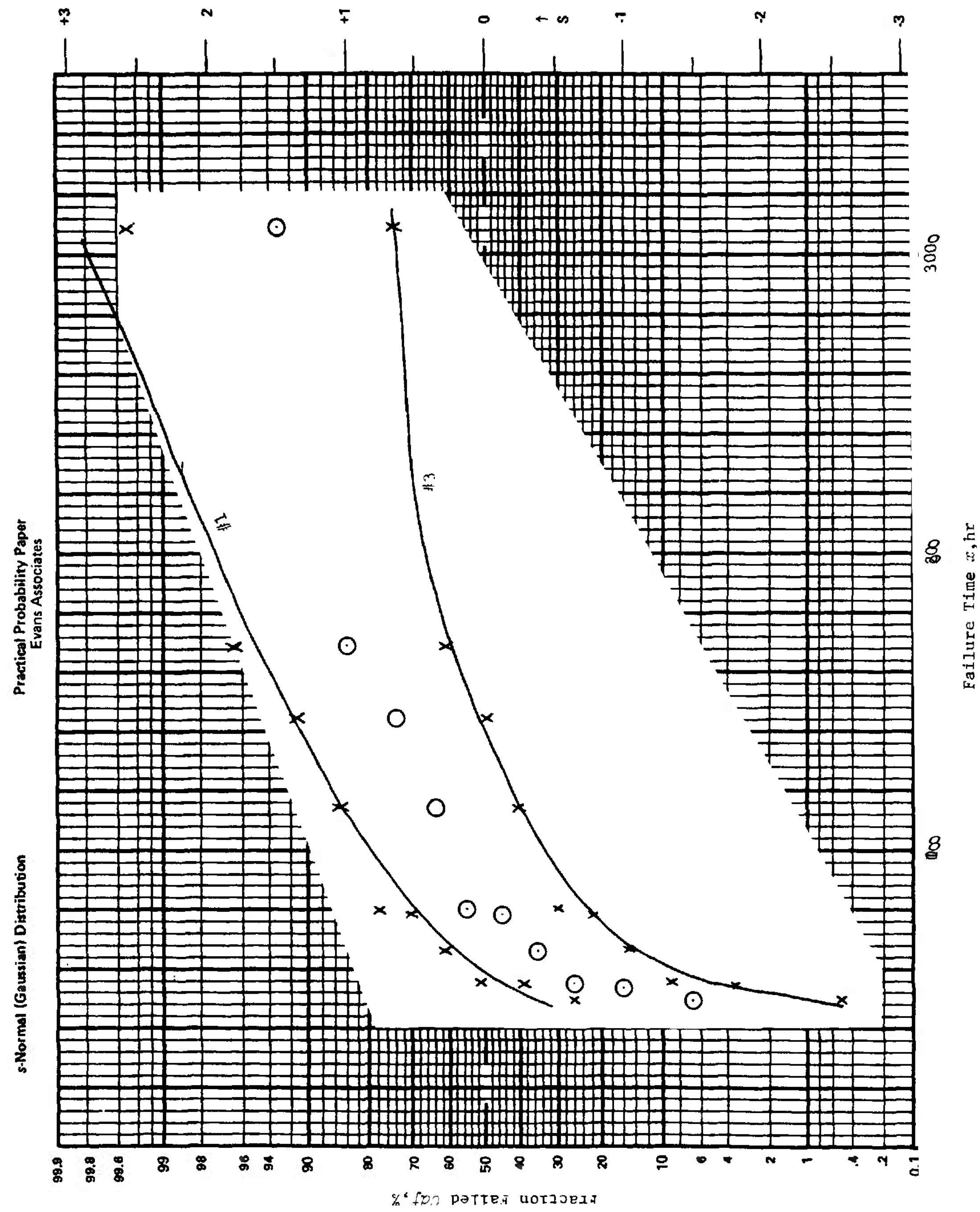


Figure 2-3 (B). Data Set C, s-Normal Scale. Beta Bounds Method

were line #3, there could be an appreciable number of early failures, although the picture doesn't *look* as bad as it did on s-normal paper (Fig. 2-1(A)).

Data Sets B and C, from Tables 2-4 and 2-5 are plotted in Fig. 2-5; the K-S Bounds method is used, with 90%s-confidence bounds (N = 10, KS, = 0.37). Data Set B is the lower half of the plotted points (circles); Data Set C is the entire set of points. For Data Set B, with only half the points, the uncertainty in the distribution appears tremendous. The lines are drawn for Data Set C, the full sample. Following the instructions on the reverse side of Fig. 2-5, one obtains

$$\alpha_L = 700 \text{ hr}, \ \hat{\alpha} = 1150 \text{ hr}, \ \alpha_U = 1500 \text{ hr}$$

$$\beta_L = 0.55, \ \hat{\beta} = 2.0, \ \beta_U = 3.6$$

This sample could easily have come from a Weibull distribution. The characteristic life (α)is known to within a factor of 2; the B_{10} life (Cdf = 10%)is probably between 10 hr and 500 hr.

Data Set D (Table 2-6) was simulated from a table of pseudo-random numbers. The procedure for plotting it on Weibull probability paper using the K-S Bounds method is shown in some detail; see Example No. 1(F).

2-2.5 LOGNORMAL DISTRIBUTION

There are several forms for writing the lognormal distribution—some of them are confusing because of a carryover from the s-normal distribution of the (μ,σ) notation. One never quite knows whether μ and σ refer to mean and standard deviation at all, and if they do, whether it is to the logs or not. The form used here is:

$$\text{lognormal } Cdf\{t\,;\alpha,\beta\}$$
$$= gauf\ [\ln((t/\alpha)^{\beta})] \quad (2\text{-}13)$$

where

α = scale parameter

β = shape parameter

The distribution is discussed more fully in par. 2-3.6 and in *Part Six, Mathematical Appendix and Glossary.*

Figs. 2-7, 2-8, and 2-9 show Data Sets A,
[text continues on page 2-36]

TABLE 2-6

DATA SET D

This is a simulated data set from random-number tables. N = 21.

Rank	Time to failure, hr	r/N	Rotting Positions, % (interpolatedfrom Table 2-1(B) between N = 20 and N = 22)
1	81.6	4.8	0.25/3.3/13
2	90.8	9.5	1.7/7.9/21
3	107	14.3	4.1/13/28
4	118	19.0	6.9/17/34
5	135	23.8	10/22/39
6	141	28.6	14/27/44
7	152	33.3	17/31/49
8	161	38.1	20/36/54
9	162	42.9	23/41/59
10	181	47.6	27/45/64
11	206	52.4	31/50/69
12	206	57.1	36/55/73
13	234	61.9	41/59/77
14	240	66.7	46/64/80
15	244	71.4	51/69/83
16	245	76.2	56/73/86
17	247	81.0	61/78/90
18	261	85.7	66/83/93.1
19	279	90.5	72/87/95.9
20	279	95.2	79/92.1/98.3
21	281	100.0	87/96.7/99.75

Column 4 is shown only for completeness; it is not illustrated with a figure.
1 − 1/N = 0.952; use 95%, s-confidence. KS_n = 0.29

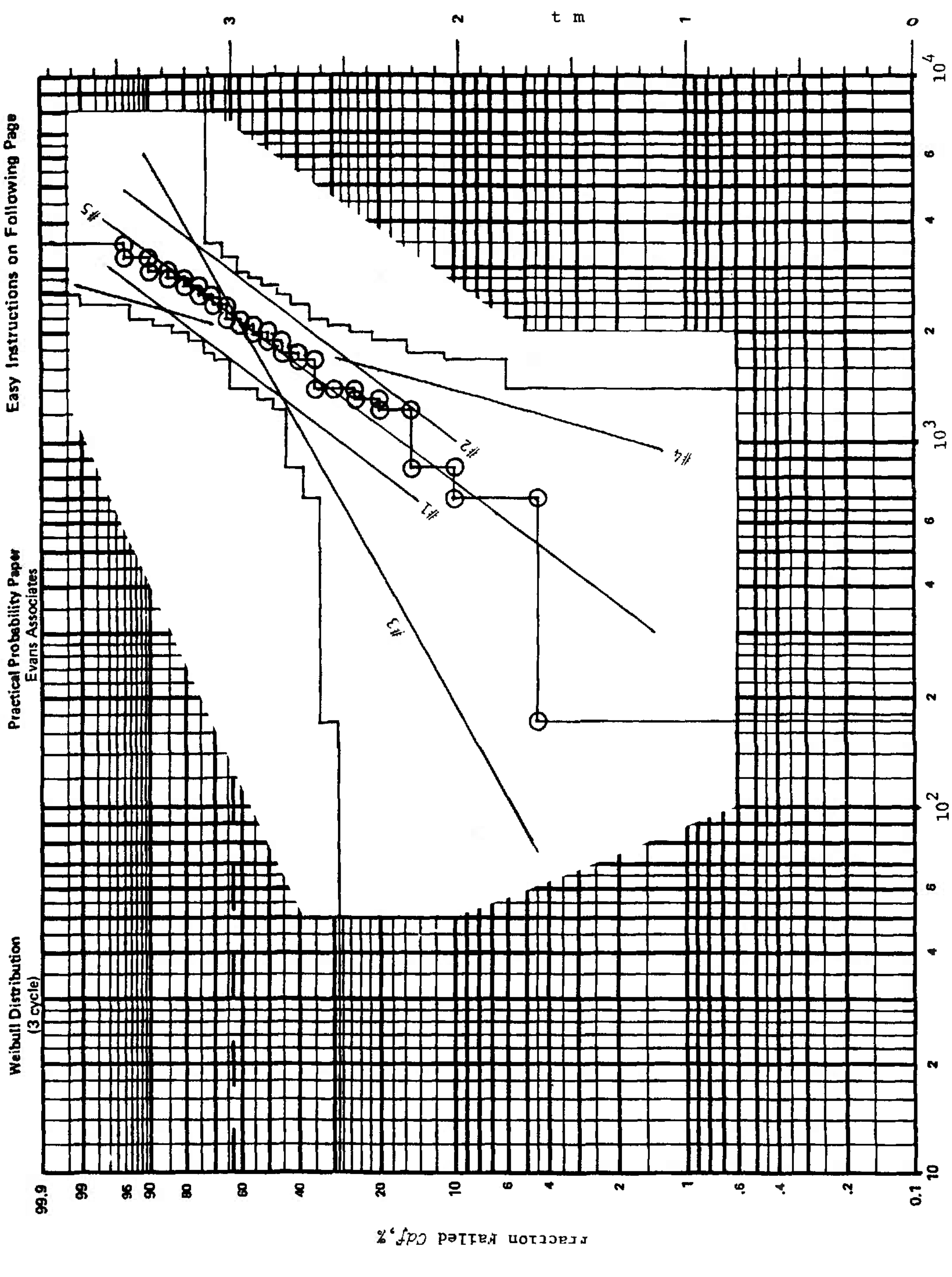


Figure 2-4. Data Set A, Weibull Scale. K-S Bounds Method (95% s-Confidence Bounds)

Practical Probability Paper
Evans Associates

Instructions for Use

Weibull Distribution

- Plotting data: Plot failure r at the two points: $F_{Hi} = r/n$ and $F_{Lo} = (r-1)/n$. Connect the points with horizontal and vertical lines; this is the sample Cdf.

- Plotting cumulative-hazard data: Plot failure r at the two points $F_{Hi} = 1 - \exp(-H_r)$ and $F_{Lo} = 1 - \exp(-H_{r-1})$ to convert the sample cumulative-hazard H_r to the Cdf. Connect the points with horizontal and vertical lines; this is the sample Cdf. The s-confidence bounds will not be exact.

- 2-sided s-confidence bounds on the actual Cdf: Choose the s-confidence level, near $1-(1/n)$ is reasonable; then find KS_n from the body of the Table (e.g., n=10, *s*-conf=**95%**, KS_n=0.41). The upper bound is plotted at $F_{Lo}+KS_n$ and $F_{Hi}+KS_n$; the lower bound is plotted at $F_{Hi}-KS_n$ and $F_{Lo}-KS_n$. For each bound, connect the points with horizontal and vertical lines; they will be parallel to, and KS_n from, the sample Cdf. Then $1-s$-conf is the fraction of times you-do-this-procedure that the true Cdf will partly lie outside the 2-sided s-confidence bounds. In general, you will be disheartened at how little you know about the true Cdf.

- Drawing data-lines: Draw the two parallel lines, farthest apart, that fit reasonably within the s-confidence bounds; use both to estimate bounds on a. Draw the two intersecting lines, with steepest and smallest slopes, that fit reasonably within the s-confidence bounds; use both to estimate bounds on b.

- To estimate a: a is the value of x at which a data-line intersects the heavy dashed line (F_a = 63.2%).

- To estimate b: Take two values of x, x_1 and x_2, 1 decade apart; then find the two B values, B_1 and B_2, which correspond to x_1 and x_2, via a data-line. $b = |B_2 - B_1|$. If x_1 and x_2 are d decades apart, then use $b = |B_2 - B_1|/d$.

Table of K-S Bounds

n	KS_n (sconfidence) (90%)	(95%)	(98%)	(99%)
5	.51	.56	.63	.67
6	.47	.52	.58	.62
8	.41	.45	.51	.54
10	.37	.41	.46	.49
12	.34	.38	.42	.45
14	.31	.35	.39	.42
16	.30	.33	.37	.39
18	.28	.31	.35	.37
20	.26	.29	.33	.36
30	.22	.24	.27	.29
40	.19	.21	.24	.25
n	$\frac{1.22}{\sqrt{n+1}}$	$\frac{1.36}{\sqrt{n+1}}$	$\frac{1.52}{\sqrt{n+1}}$	$\frac{1.63}{\sqrt{n+1}}$

(formula is o.k. for $n \geqslant 6$)

For censored samples, use an n which is between the original-number-on-test and the number-of-failures.

Weibull Distribution

$F(x) = 1 - \exp[-(x/a)^b]$

F Cumulative distribution function (Cdf)
a scale parameter (same units as x)
b shape parameter (no units)
n sample size
r failure number; r = 1, 2, ... ,n

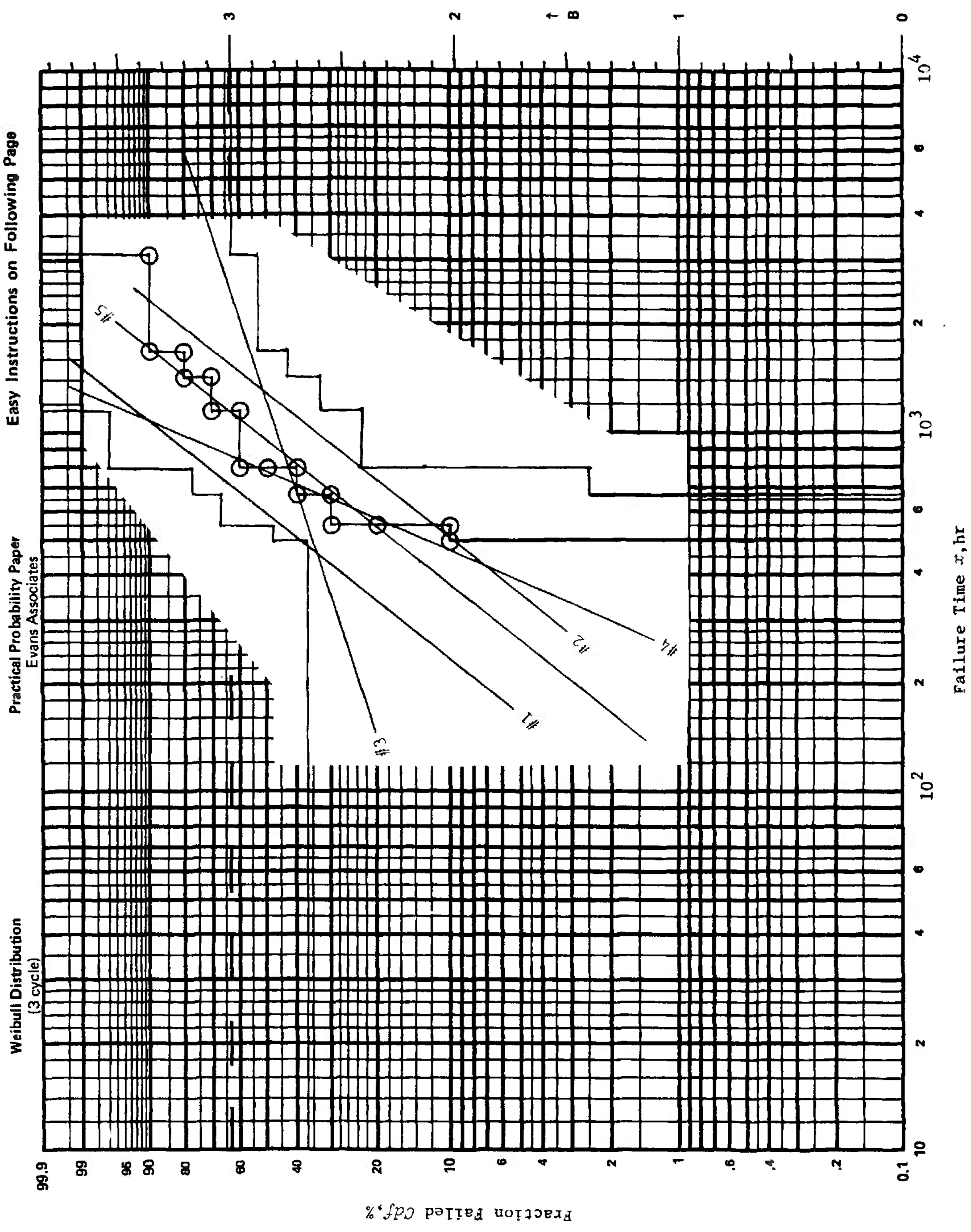


Figure 2-5. Data Sets B and C, Weibull Scale. K-S Bounds Method (90% s-Confidence Bounds)

Practical Probability Paper
Evans Associates

Instructions for Use

Weibull Distribution

- Plotting data: Plot failure r at the two points: $F_{Hi} = r/n$ and $F_{Lo} = (r-1)/n$. Connect the points with horizontal and vertical lines; this is the sample Cdf.

- Plotting cumulative-hazard data: Plot failure r at the two points $F_{Hi} = 1 - \exp(-H_r)$ and $F_{Lo} = 1 - \exp(-H_{r-1})$ to convert the sample cumulative-hazard H_r to the Cdf. Connect the points with horizontal and vertical lines; this is the sample Cdf. The s-confidence bounds will not be exact.

- 2-sided s-confidence bounds on the actual Cdf: Choose the s-confidence level, near $1-(1/n)$ is reasonable; then find KS_n from the body of the Table (e.g., n=10, s-conf=95%, KS_n=0.41). The upper bound is plotted at $F_{Lo}+KS_n$ and $F_{Hi}+KS_n$; the lower bound is plotted at $F_{Hi}-KS_n$ and $F_{Lo}-KS_n$. For each bound, connect the points with horizontal and vertical lines; they will be parallel to, and KS_n from, the sample Cdf. Then $1-s$-conf is the fraction of times you-do-this-procedure that the true Cdf will partly lie outside the 2-sided s-confidence bounds. In general, you will be disheartened at how little you know about the true Cdf.

- Drawing data-lines: Draw the two parallel lines, farthest apart, that fit reasonably within the s-confidence bounds; use both to estimate bounds on a. Draw the two intersecting lines, with steepest and smallest slopes, that fit reasonably within the s-confidence bounds; use both to estimate bounds on b.

- To estimate a: a is the value of x at which a data-line intersects the heavy dashed line (F_a = 63.2%).

- To estimate b: Take two values of x, x_1 and x_2, 1 decade apart; then find the two B values, B_1 and B_2, which correspond to x_1 and x_2, via a data-line. $b = |B_2 - B_1|$. If x_1 and x_2 are d decades apart, then use $b = |B_2 - B_1|/d$.

Table of K S Bounds

n	KS_n (sconfidence) (90%)	(95%)	(98%)	(99%)
5	.51	.56	.63	.67
6	.47	.52	.58	.62
8	.41	.45	.51	.54
10	.37	.41	.46	.49
12	.34	.38	.42	.45
14	.31	.35	.39	.42
16	.30	.33	.37	.39
18	.28	.31	.35	.37
20	.26	.29	.33	.36
30	.22	.24	.27	.29
40	.19	.21	.24	.25
n	$\frac{1.22}{\sqrt{n+1}}$	$\frac{1.36}{\sqrt{n+1}}$	$\frac{1.52}{\sqrt{n+1}}$	$\frac{1.63}{\sqrt{n+1}}$

(formula is o.k. for $n \geqslant 6$)

For censored samples, use an n which is between the original-number-on-test and the number-of-failures.

Weibull Distribution

$F(x) = 1 - \exp[-(x/a)^b]$

F Cumulative distribution function (Cdf)
a scale parameter (same units as x)
b shape parameter (no units)
n sample size
r failure number; $r = 1, 2, \ldots, n$

Example No. 1(F)

This example uses Data Set D, Table 2–6.

Data Set D, K–S Bounds Method (Table 2–6, Fig. 2–6)

Procedure	Example
1. Prepare Col. 3. Choose an s-confidence level, and find KS_n from the reverse side of Fig. 2–6.	1. $N = 21$; $1 - 1/N = 0.952$. Choose 95% s-confidence. $KS_n = 0.29$ (from the formula).
2. Plot the data in Col. 3 and the bounds, as indicated on the reverse side of Fig. 2–6.	2. See Fig. 2–6.
3. Draw the two parallel lines for estimating the lower and upper limits of α, α_L and α_U. α_L and α_U are the intersections of lines #1 and #2 with the heavy dashed line.	3. See lines #1 and #2. $\alpha_L = 170$ hr $\alpha_U = 280$ hr.
4. Draw the two intersecting lines for estimating the lower and upper limits of β, β_L and β_U. β_L and β_U are calculated from the slopes of lines #3 and #4.	4. For line #3, choose the decade from 30 to 300. $B_1 = 2.05$, $B_2 = 3.10$; thus $\beta_L = 3.10 - 2.05 = 1.05$. For line #4, choose the half-decade from 100 to 316. $B_1 = 1.10$, $B_2 = 3.90$; thus $\beta_U = (3.90 - 1.10)/0.5 = 5.6$.
5. Use the "best fit" line to make point estimates of α and β, $\hat{\alpha}$ and $\hat{\beta}$.	5. From line #5: $\hat{\alpha} = 220$ hr $\hat{\beta} = 3.0$.

These data could easily have come from a Weibull distribution, although the data do not lie on a very straight line. The parent population actually was Weibull, with $\alpha = 200$ hr and $\beta = 2.5$. This line (not shown in Fig. 2–6) is well within the K–S envelope, but is not a good fit to the data points. The data themselves do not lie well on their best-fit line (#5). Be very skeptical when the data do lie on a nice straight line; someone may have given the data a preliminary manipulation.

Example No. 1(F) (Cont'd)

Do not use graphical methods to estimate a location parameter for the Weibull distribution; you will only be playing a losing game with random numbers. If you must estimate a location parameter for the Weibull distribution, consult a competent statistician. The arithmetic manipulations can easily be done precisely (on a computer); but it is so easy to mislead oneself.

The big lesson to be learned from Data Set D is that a random sample of about 20 points or less can only give ballpark estimates -- you learn the name of the game and where it's being played, that's all.

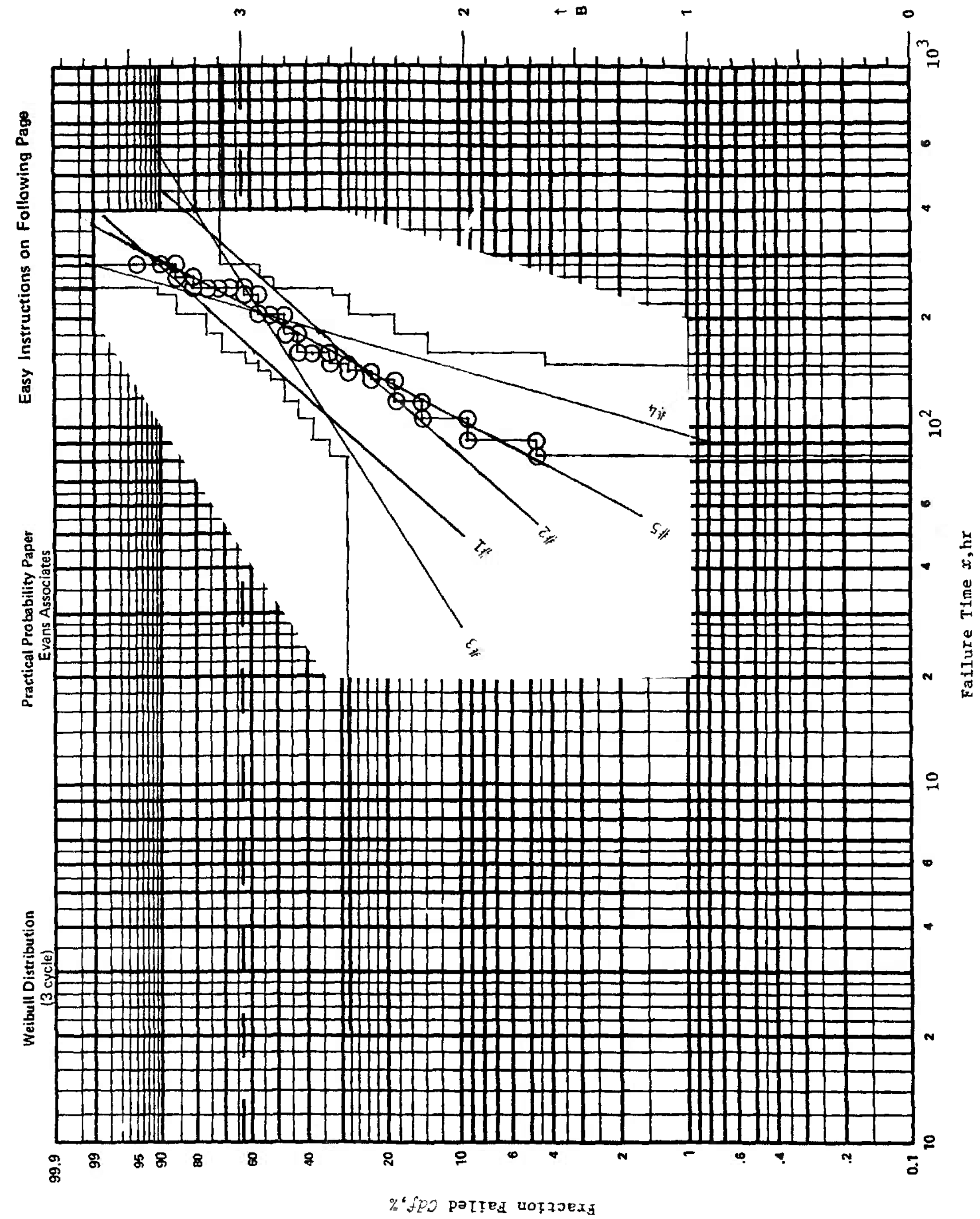


Figure 2-6. Data Set D, Weibull Scale. K-S Bounds Method (95% s-Confidence Bounds)

Practical Probability Paper
Evans Associates

Instructions for Use

Weibull Distribution

- Plotting data: Plot failure r at the two points: $F_{Hi} = r/n$ and $F_{Lo} = (r-1)/n$. Connect the points with horizontal and vertical lines; this is the sample Cdf.

- Plotting cumulative-hazard data: Plot failure r at the two points $F_{Hi} = 1 - \exp(-H_r)$ and $F_{Lo} = 1 - \exp(-H_{r-1})$ to convert the sample cumulative-hazard H_r to the Cdf. Connect the points with horizontal and vertical lines; this is the sample Cdf. The s-confidence bounds will not be exact.

- 2-sided s-confidence bounds on the actual Cdf: Choose the s-confidence level, near $1-(1/n)$ is reasonable; then find KS_n from the body of the Table (e.g., n=10, s-conf=95%, KS_n=0.41). The upper bound is plotted at $F_{Lo}+KS_n$ and $F_{Hi}+KS_n$; the lower bound is plotted at $F_{Hi}-KS_n$ and $F_{Lo}-KS_n$. For each bound, connect the points with horizontal and vertical lines; they will be parallel to, and KS_n from, the sample Cdf. Then 1 − s-conf is the fraction of times you-do-this-procedure that the true Cdf will partly lie outside the 2-sided sconfidence bounds. In general, you will be disheartened at how little you know about the true Cdf.

- Drawing data-lines: Draw the two parallel lines, farthest apart, that fit reasonably within the s-confidence bounds; use both to estimate bounds on a. Draw the two intersecting lines, with steepest and smallest slopes, that fit reasonably within the s-confidence bounds; use both to estimate bounds on b.

- To estimate a: a is the value of x at which a data-line intersects the heavy dashed line (F_a = 63.2%).

- To estimate b: Take two values of x, x_1 and x_2, 1 decade apart; then find the two B values, B_1 and B_2, which correspond to x_1 and x_2, via a data-line. $b = |B_2 - B_1|$. If x_1 and x_2 are d decades apart, then use $b = |B_2 - B_1|/d$.

Table of K-S Bounds

n	KS_n (sconfidence) (90%)	(95%)	(98%)	(99%)
5	.51	.56	.63	.67
6	.47	.52	.58	.62
8	.41	.45	.51	.54
10	.37	.41	.46	.49
12	.34	.38	.42	.45
14	.31	.35	.39	.42
16	.30	.33	.37	.39
18	.28	.31	.35	.37
20	.26	.29	.33	.36
30	.22	.24	.27	.29
40	.19	.21	.24	.25
n	$\frac{1.22}{\sqrt{n+1}}$	$\frac{1.36}{\sqrt{n+1}}$	$\frac{1.52}{\sqrt{n+1}}$	$\frac{1.63}{\sqrt{n+1}}$

(formula is o.k. for $n \geqslant 6$)

For censored samples, use an n which is between the original-number-on-test and the number-of-failures.

Weibull Distribution

$F(x) = 1 - \exp[-(x/a)^b]$

F Cumulative distribution function (Cdf)
a scale parameter (same units as x)
b shape parameter (no units)
n sample size
r failure number; r = 1, 2, ... , n

C, and D plotted, respectively, on lognormal probability paper using the K-S Bounds method. The s-confidence levels are the same as used for the same data sets in earlier paragraphs. See instructions for use on the reverse side of the Practical Probability Paper.

Data Set A (Fig. 2-7) could reasonably have come from a lognormal distribution. The parameter estimates, from lines #1–#5 are:

$$\alpha_L = 1450 \text{ hr}, \; \hat{\alpha} = 1900 \text{ hr},$$

$$\alpha_U = 2400 \text{ hr}$$

$$\beta_L = 4.0, \; \hat{\beta} = 11, \; \beta_U = 27 .$$

Data Set C (Fig. 2-8) could reasonably have come from a lognormal distribution. No parameter estimates were made because they would serve little purpose. The reason Data Set *C* could reasonably have come from a lognormal distribution is that a variety of straight lines could be drawn within the K-S envelope in Fig. 2-8. Anyone of those lines could be the actual distribution.

Data Set D (Fig. 2-9) could reasonably have come from a lognormal distribution. No parameter estimates were made.

Small samples (e.g., 20 or less) cannot pinpoint the distribution from whence they came. Lognormal and Weibull distributions are difficult to tell apart even with somewhat larger samples.

2-2.6 SUMMARY

Small samples (say, less than 20) are notoriously unreliable indicators for the shape of a population. Always plot the K-S bounds along with the sample *Cdf*. Figs. 2-10 and 2-11 show the behavior of samples from the uniform distribution (see Ref. 27). Fig. 2-10 shows the first 10 order statistics out of a sample of 99 (the plots are vs the mean plotting value from Eq. 2-4). The horizontal and vertical scales are the same; so the population (shown solid) plots as a 45-deg straight line. Fig. 2-11 shows all 9 order statistics from a sample of 9. Again, the horizontal and vertical scales are the same; so the population plots as a 45-deg straight line. (The K-S bounds are not shown here in this special case, so as not to clutter up the figures.)

All 3 data sets (A,C,D) could reasonably have come from any of the 3 distributions: s-normal, Weibull, or lognormal. It is fruitless to try to pick the "best" one; that would be an exercise in sterile, pointless mathematics.

Before you pursue the questions of which distributions and what parameter values, ask yourself "Why do I want to know?" If it's because you want to estimate some property of the population, stop. You're kidding yourself if you think mathematical manipulations will tell you the answer when the graph says you don't know.

Do not try to estimate a location parameter for Weibull and lognormal distributions by trying to "straighten-out" the sample plot. If you are ever tempted to do it, look again at Figs. 2-10 and 2-11 which show how "untypical" a random sample usually is.

Graphical estimation usually can provide all the accuracy that the data can use. It can give a good picture of the uncertainty in any extrapolation (or interpolation) from the sample. It can also make you think harder about why you wanted to estimate some parameters from the data. When an iterative analytic method is used to estimate parameters, the graphical estimates are usually excellent choices for the starting point.

Beware of data from small samples that plot as very straight lines on any kind of probability paper. The chances of its hap-

[text continues on page 2-49]

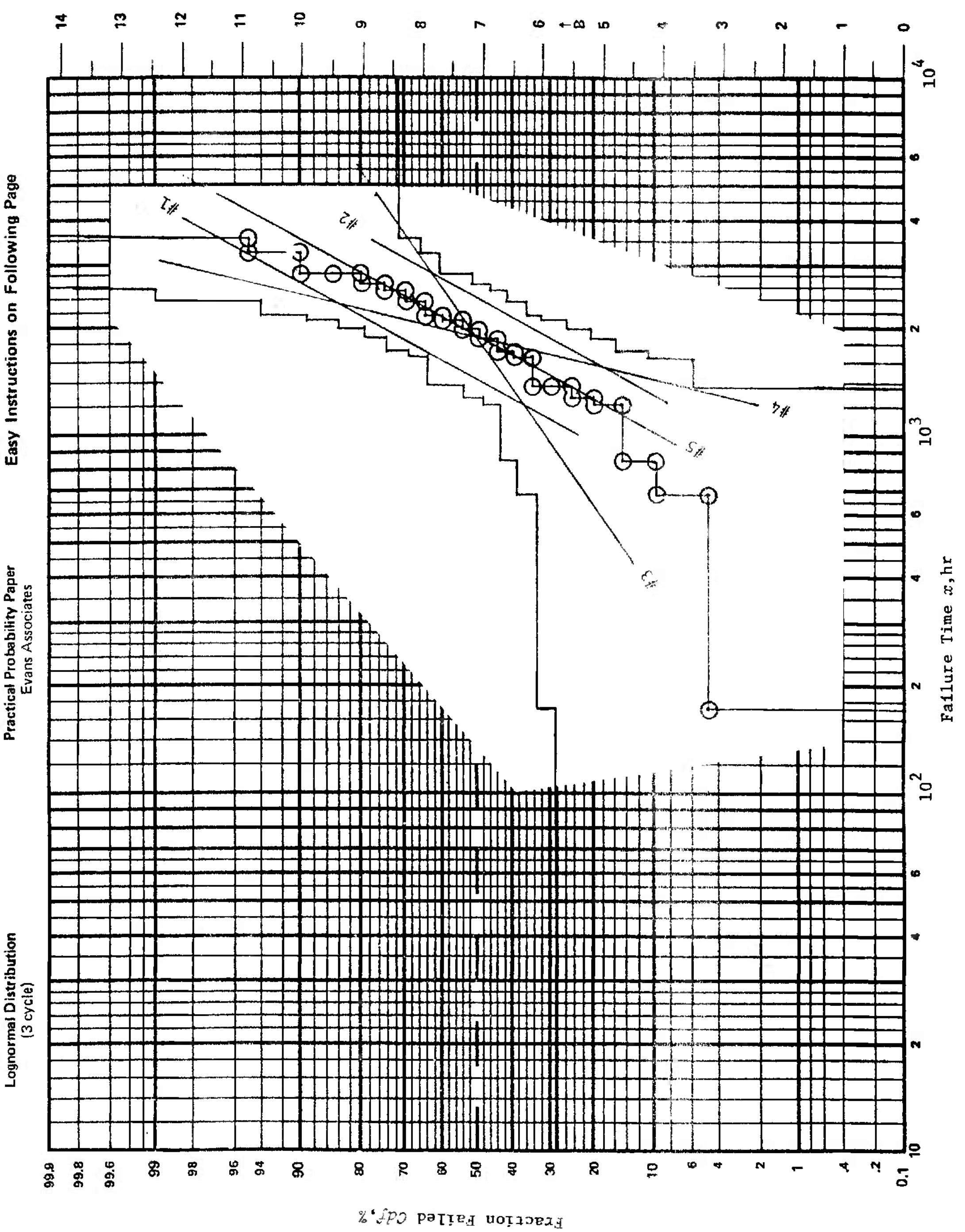


Figure 2-7. Data Set A, Lognormal Scale. K-S Bounds Method (95% s-Confidence)

Instructions for Use

Lognormal Distribution

- Plotting data: Plot failure r at the two points: $F_{Hi} = r/n$ and $F_{Lo} = (r-1)/n$. Connect the points with horizontal and vertical lines; this is the sample Cdf.

- Plotting cumulative-hazard data: Plot failure r at the two points $F_{Hi} = 1 - \exp(-H_r)$ and $F_{Lo} = 1 - \exp(-H_{r-1})$ to convert the sample cumulative-hazard H_r to the Cdf. Connect the points with horizontal and vertical lines; this is the sample Cdf. The sconfidence bounds will not be exact.

- 2-sided s-confidence bounds on the actual Cdf: Choose the s-confidence level, near $1-(1/n)$ is reasonable; then find KS_n from the body of the Table (e.g., n=10, s-conf=95%, KS_n=0.41). The upper bound is plotted at $F_{Lo}+KS_n$ and $F_{Hi}+KS_n$; the lower bound is plotted at $F_{Hi}-KS_n$ and $F_{Lo}-KS_n$. For each bound, connect the points with horizontal and vertical lines; they will be parallel to, and KS_n from, the sample Cdf. Then $1-s$-conf is the fraction of times you-do-this-procedure that the true Cdf will partly lie outside the 2-sided s-confidence bounds. In general, you will be disheartened at how little you know about the true Cdf.

- Drawing data-lines: Draw the two parallel lines, farthest apart, that fit reasonably within the s-confidence bounds; use both to estimate bounds on a. Draw the two intersecting lines, with steepest and smallest slopes, that fit reasonably within the s-confidence bounds; use both to estimate bounds on b.

- To estimate a: a is the value of x at which a data-line intersects the heavy dashed line (F_a = 50.0%).

- To estimate b: Take two values of x, x_1 and x_2, 1 decade apart; then find the two B values, B_1 and B_2, which correspond to x_1 and x_2, via a data-line. $b = |B_2 - B_1|$. If x_1 and x_2 are d decades apart, then use $b = |B_2 - B_1|/d$.

Table of K-S Bounds

n	KS_n (sconfidence) (90%)	(95%)	(98%)	(99%)
5	.51	.56	.63	.67
6	.47	.52	.58	.62
8	.41	.45	.51	.54
10	.37	.41	.46	.49
12	.34	.38	.42	.45
14	.31	.35	.39	.42
16	.30	.33	.37	.39
18	.28	.31	.35	.37
20	.26	.29	.33	.36
30	.22	.24	.27	.29
40	.19	.21	.24	.25
n	$\frac{1.22}{\sqrt{n+1}}$	$\frac{1.36}{\sqrt{n+1}}$	$\frac{1.52}{\sqrt{n+1}}$	$\frac{1.63}{\sqrt{n+1}}$

(formula is o.k. for $n \geqslant 6$)

For censored samples, use an n which is between the original-number-on-test and the number-of-failures.

Lognormal Distribution

$F(x) = \text{gauf}\,(\log_e[(\frac{x}{a})^b])$, gauf (z) is the standard s-normal Cdf

F Cumulative distribution function (Cdf)

a scale parameter (same units as x), also the median; log a is the median and mean (average) of $\log_e x$

b shape parameter (dimensionless); $1/b$ is the standard deviation of $\log_e x$

n sample size

r failure number; r = 1, 2, ... , n

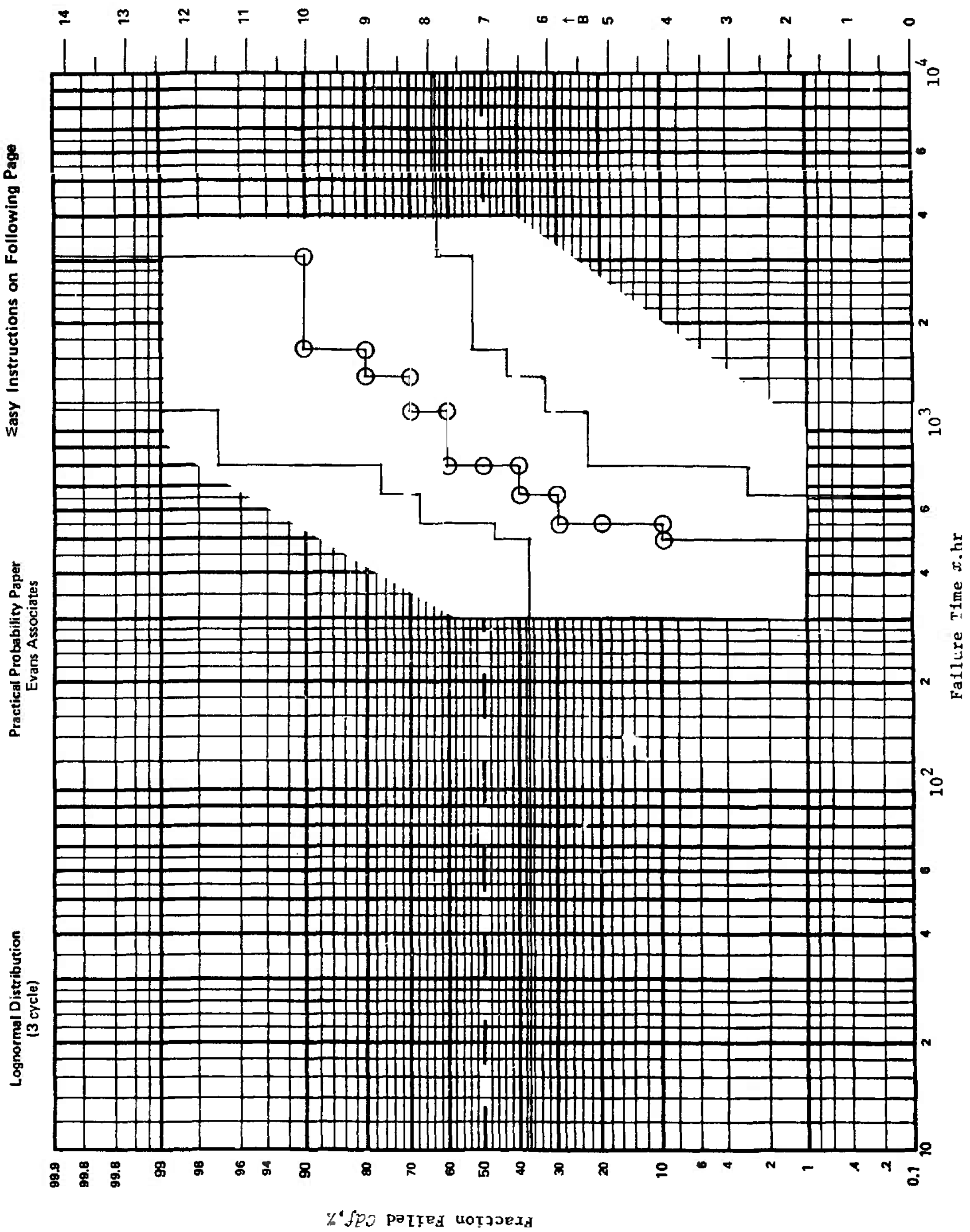


Figure 2-8. [illegible] Set [illegible]. Lognormal Scale. K-S Bounds Method (90% s-Confidence)

Practical Probability Paper
Evans Associates

Instructions for Use

Lognormal Distribution

- Plotting data: Plot failure r at the two points: $F_{Hi} = r/n$ and $F_{Lo} = (r-1)/n$. Connect the points with horizontal and vertical lines; this is the sample Cdf.

- Plotting cumulative-hazard data: Plot failure r at the two points $F_{Hi} = 1 - \exp(-H_r)$ and $F_{Lo} = 1 - \exp(-H_{r-1})$ to convert the sample cumulative-hazard H_r to the Cdf. Connect the points with horizontal and vertical lines; this is the sample Cdf. The s-confidence bounds will not be exact.

- 2-sided s-confidence bounds on the actual Cdf: Choose the s-confidence level, near $1-(1/n)$ is reasonable; then find KS_n from the body of the Table (e.g., n=10, s-conf=95%, KS_n=0.41). The upper bound is plotted at $F_{Lo}+KS_n$ and $F_{Hi}+KS_n$; the lower bound is plotted at $F_{Hi}-KS_n$ and $F_{Lo}-KS_n$. For each bound, connect the points with horizontal and vertical lines; they will be parallel to, and KS_n from, the sample Cdf. Then 1 — s-conf is the fraction of times you-do-this-procedure that the true Cdf will partly lie outside the 2-sided s-confidence bounds. In general, you will be disheartened at how little you know about the true Cdf.

- Drawing data-lines: Draw the two parallel lines, farthest apart, that fit reasonably within the s-confidence bounds; use both to estimate bounds on ***a***. Draw the two intersecting lines, with steepest and smallest slopes, that fit reasonably within the s-confidence bounds; use both to estimate bounds on ***b***.

- To estimate a: a is the value of x at which a data-line intersects the heavy dashed line (F_a = 50.0%).

- To estimate ***b***: Take two values of x, x_1 and x_2, 1 decade apart; then find the two B values, B_1 and B_2, which correspond to x_1 and x_2, via a data-line. $b = |B_2 - B_1|$. If x_1 and x_2 are d decades apart, then use $b = |B_2 - B_1|/d$.

Table of K-S Bounds

n	KS_n (s-confidence) (90%)	(95%)	(98%)	(99%)
5	.51	.56	.63	.67
6	.47	.52	.58	.62
8	.41	.45	.51	.54
10	.37	.41	.46	.49
52	.34	.38	.42	.45
14	.31	.35	.39	.42
16	.30	.33	.37	.39
18	.28	.31	.35	.37
20	.26	.29	.33	.36
30	.22	.24	.27	.29
40	.19	.21	.24	.25
n	$\frac{1.22}{\sqrt{n+1}}$	$\frac{1.36}{\sqrt{n+1}}$	$\frac{1.52}{\sqrt{n+1}}$	$\frac{1.63}{\sqrt{n+1}}$

(formula is o.k. for $n \geqslant 6$)

For censored samples, use an n which is between the original-number-on-test and the number-of-failures.

Lognormal Distribution

$F(x)$ = gauf $(\log_e[(\frac{x}{a})^b])$, gauf (z) is the standard s-normal Cdf

F Cumulative distribution function (Cdf)

a scale parameter (same units as x), also the median; $\log a$ is the median and mean (average) of $\log_e x$

b shape parameter (dimensionless); $1/b$ is the standard deviation of $\log_e x$

n sample size

r failure number; r = 1, 2, ... , n

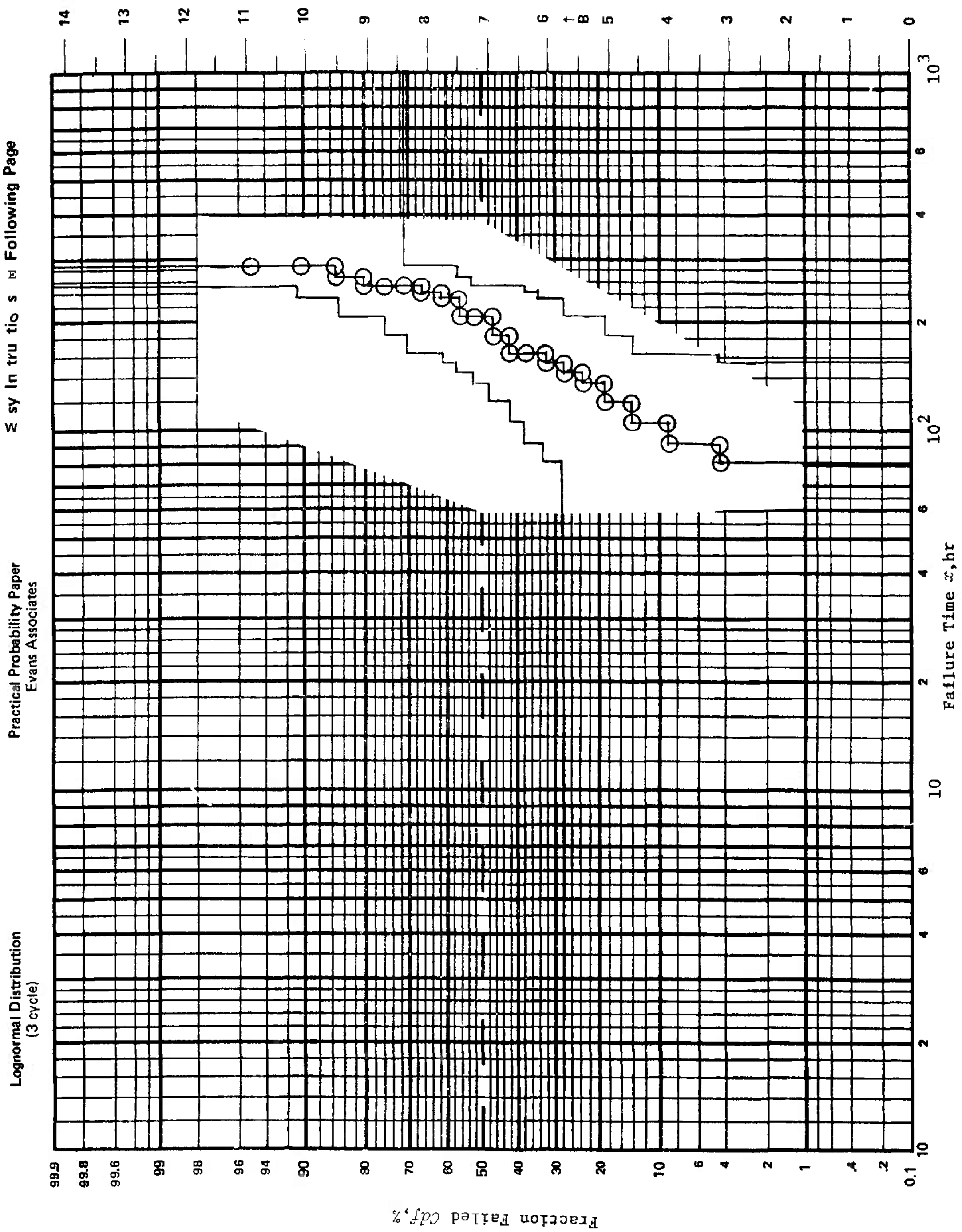


Figure 2-9. Data Set D, Lognormal Scale, 4-S Bounds (95% s-Confidence)

Instructions for Use

Lognormal Distribution

- Plotting data: Plot failure r at the two points: $F_{Hi} = r/n$ and $F_{Lo} = (r-1)/n$. Connect the points with horizontal and vertical lines; this is the sample Cdf.

- Plotting cumulative-hazard data: Plot failure r at the two points $F_{Hi} = 1 - \exp(-H_r)$ and $F_{Lo} = 1 - \exp(-H_{r-1})$ to convert the sample cumulative-hazard H_r to the Cdf. Connect the points with horizontal and vertical lines; this is the sample Cdf. The s-confidence bounds will not be exact.

- 2-sided s-confidence bounds on the actual Cdf: Choose the s-confidence level, near $1-(1/n)$ is reasonable; then find KS_n from the body of the Table (e.g., n=10, s-conf=95%, KS_n=0.41). The upper bound is plotted at $F_{Lo}+KS_n$ and $F_{Hi}+KS_n$; the lower bound is plotted at $F_{Hi}-KS_n$ and $F_{Lo}-KS_n$. For each bound, connect the points with horizontal and vertical lines; they will be parallel to, and KS_n from, the sample Cdf. Then 1 − s-conf is the fraction of times you-do-this-procedure that the true Cdf will partly lie outside the 2-sided s-confidence bounds. In general, you will be disheartened at how little you know about the true Cdf.

- Drawing data-lines: Draw the two parallel lines, farthest apart, that fit reasonably within the s-confidence bounds; use both to estimate bounds on a. Draw the two intersecting lines, with steepest and smallest slopes, that fit reasonably within the s-confidence bounds; use both to estimate bounds on b.

- To estimate a: a is the value of x at which a data-line intersects the heavy dashed line (F_a = 50.0%).

- To estimate b: Take two values of x, x_1 and x_2, 1 decade apart; then find the two B values, B_1 and B_2, which correspond to x_1 and x_2, via a data-line. $b = |B_2 - B_1|$. If x_1 and x_2 are d decades apart, then use $b = |B_2 - B_1|/d$.

Table of K-S Bounds

n	KS_n (s-confidence) (90%)	(95%)	(98%)	(99%)
5	.51	.56	.63	.67
6	.47	.52	.58	.62
8	.41	.45	.51	.54
10	.37	.41	.46	.49
12	.34	.38	.42	.45
14	.31	.35	.39	.42
16	.30	.33	.37	.39
18	.28	.31	.35	.37
20	.26	.29	.33	.36
30	.22	.24	.27	.29
40	.19	.21	.24	.25
n	$\frac{1.22}{\sqrt{n+1}}$	$\frac{1.36}{\sqrt{n+1}}$	$\frac{1.52}{\sqrt{n+1}}$	$\frac{1.63}{\sqrt{n+1}}$

(formula is o.k. for $n \geq 6$)

For censored samples, use an n which is between the original-number-on-test and the number-of-failures.

Lognormal Distribution

$F(x) = \text{gauf}\,(\log_e[(\frac{x}{a})^b])$, gauf (z) is the standard s-normal Cdf

F Cumulative distribution function (Cdf)

a scale parameter (same units as x), also the median; log a is the median and mean (average) of $\log_e x$

b shape parameter (dimensionless); $1/b$ is the standard deviation of $\log_e x$

n sample size

r failure number; r = 1, 2, ... , n

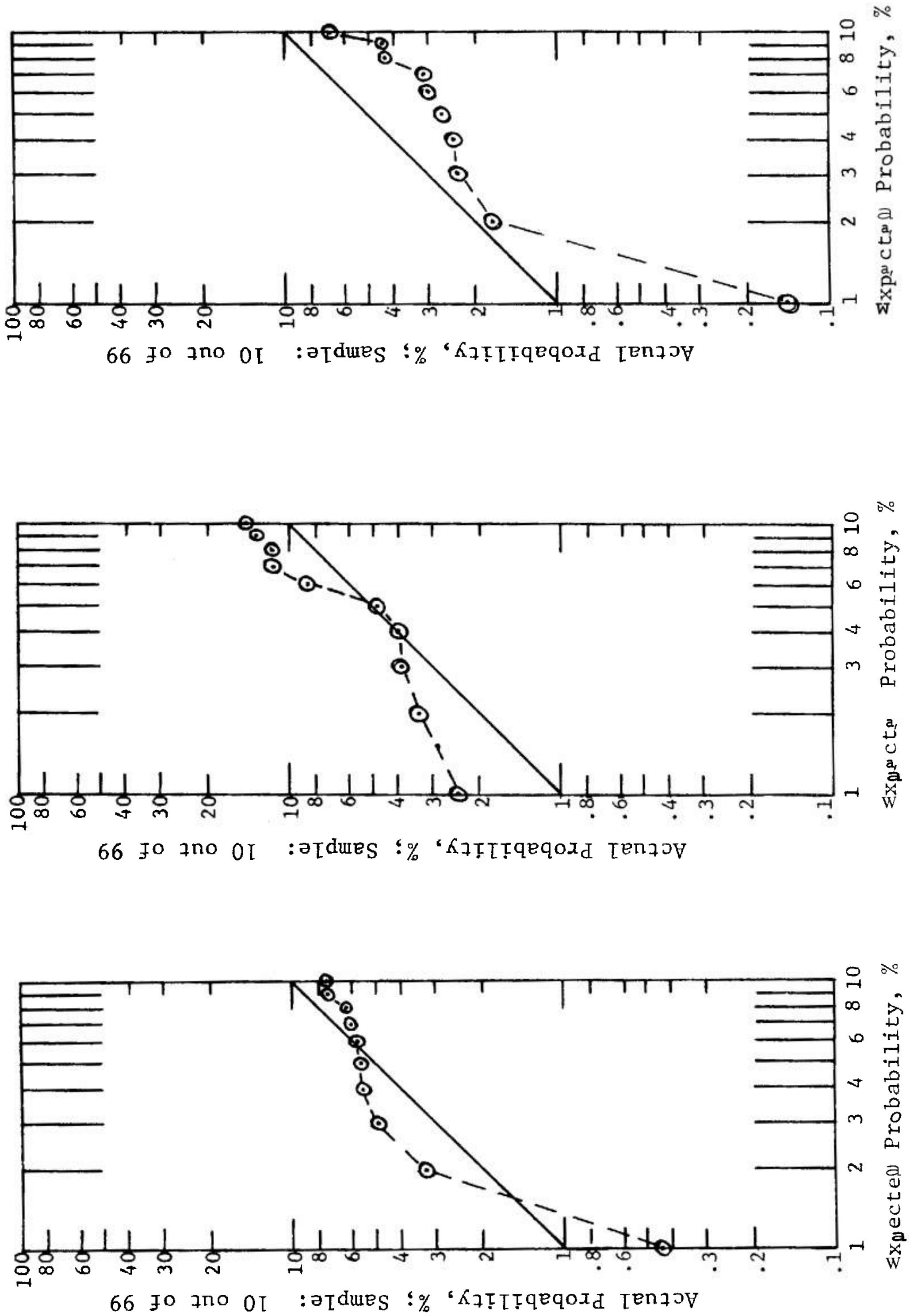


Figure 2-10. Random Samples, Smallest 10 Out of 99 (Ref. 27)

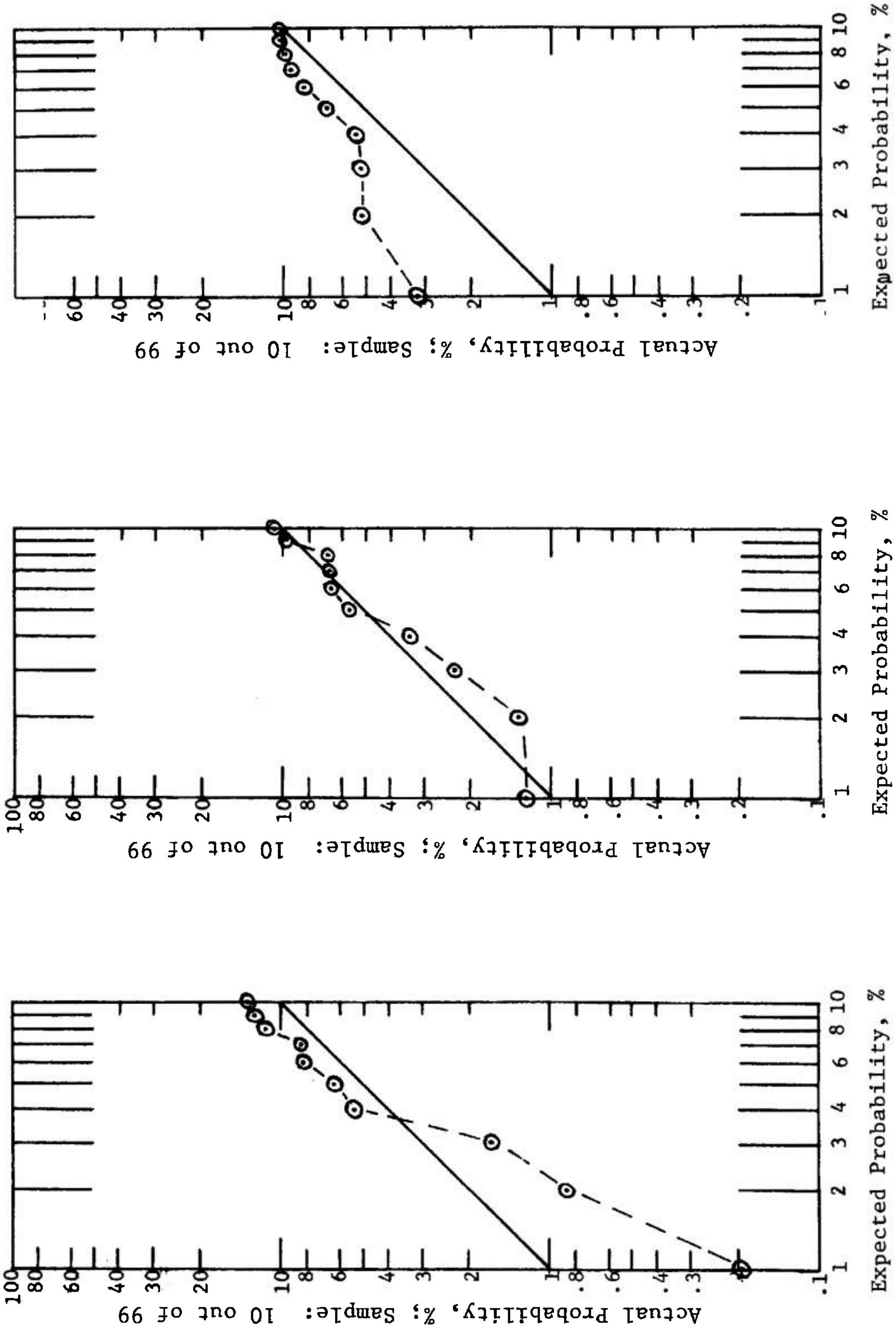


Figure 2-10. (Cont'd)

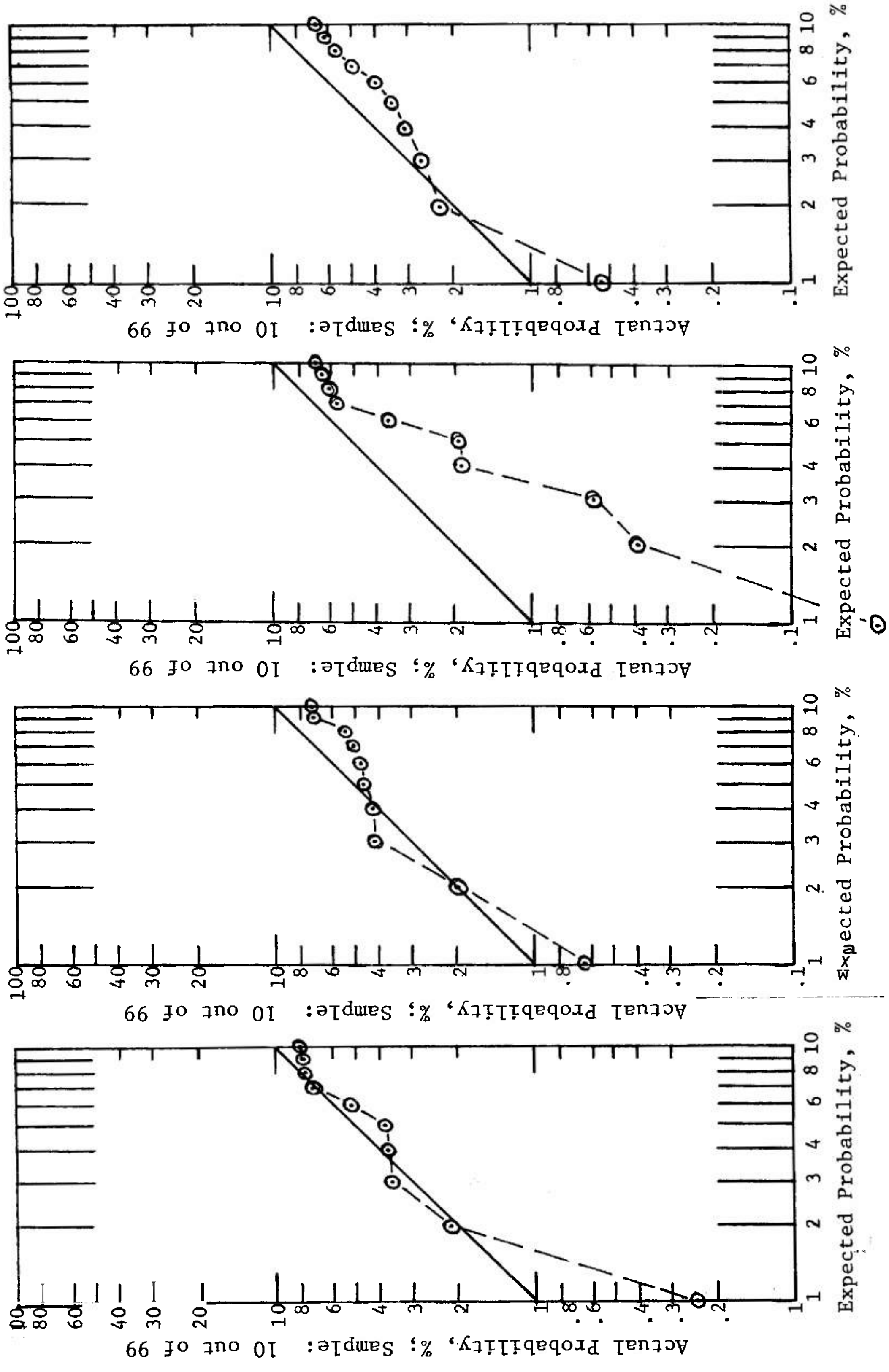


Figure 2-10. (Cont'd)

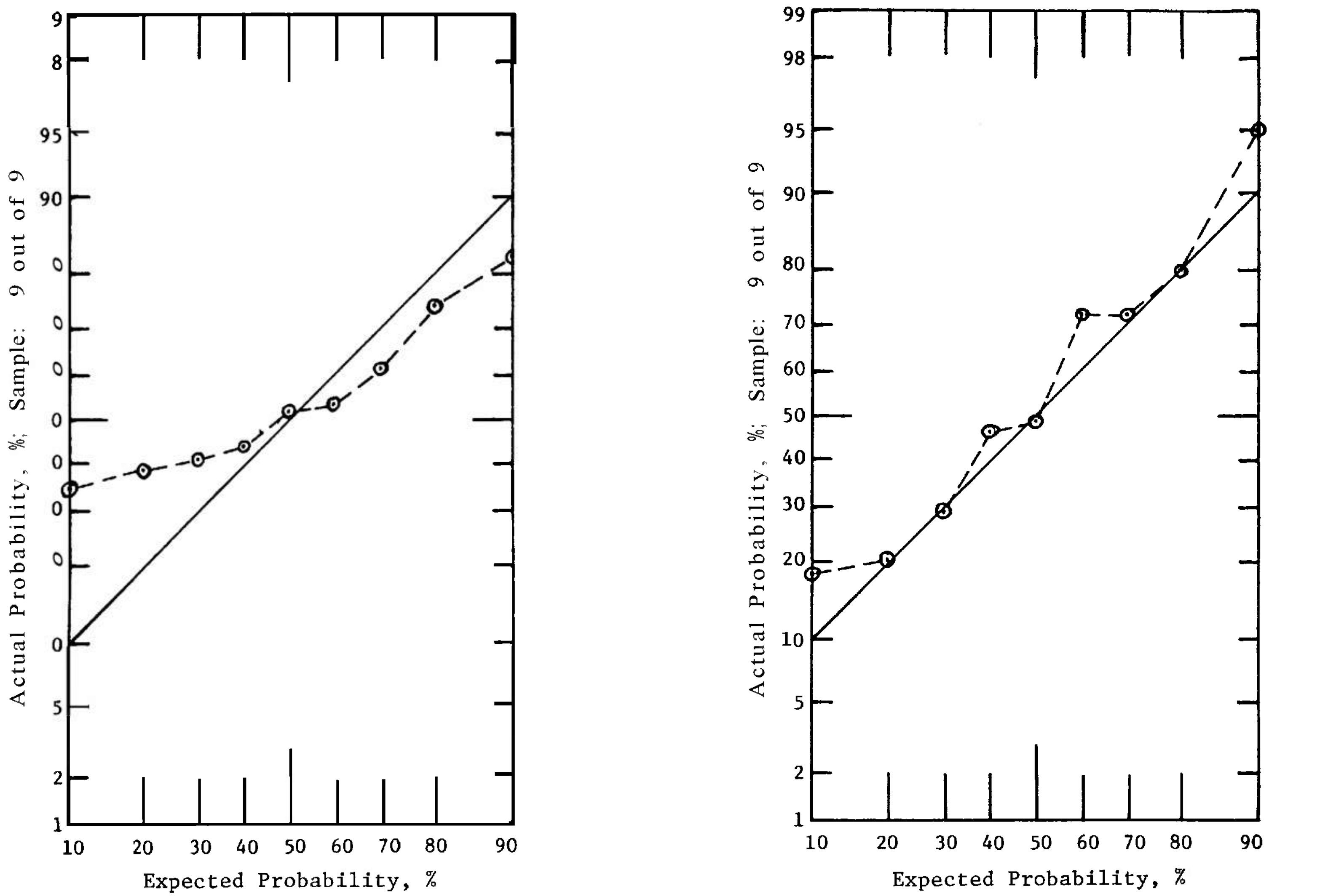
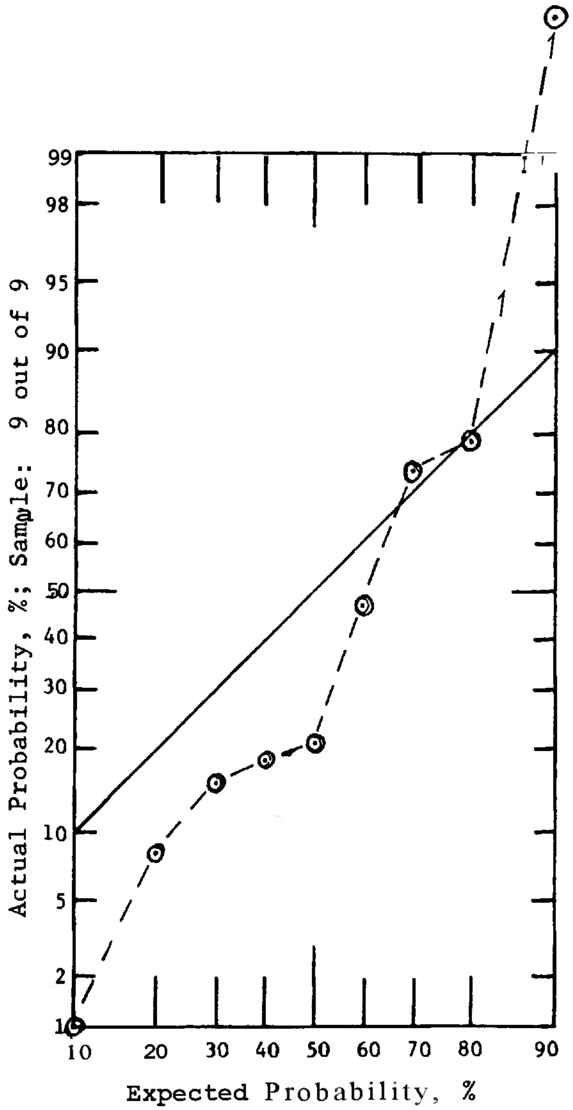


Figure 2-11. Random Samples, Complete 9 Out of 9 (Ref. 27)



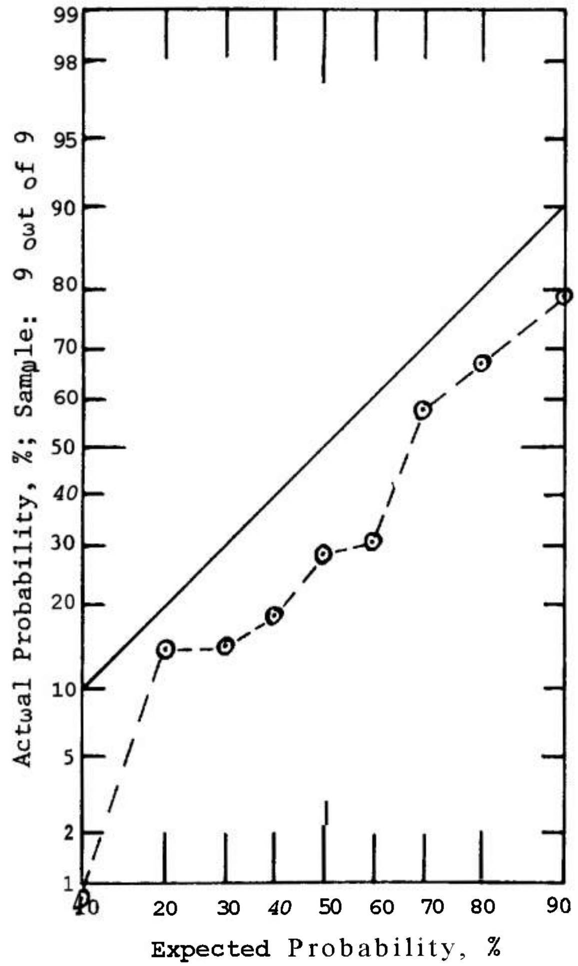


Figure 2-11. (Cont'd)

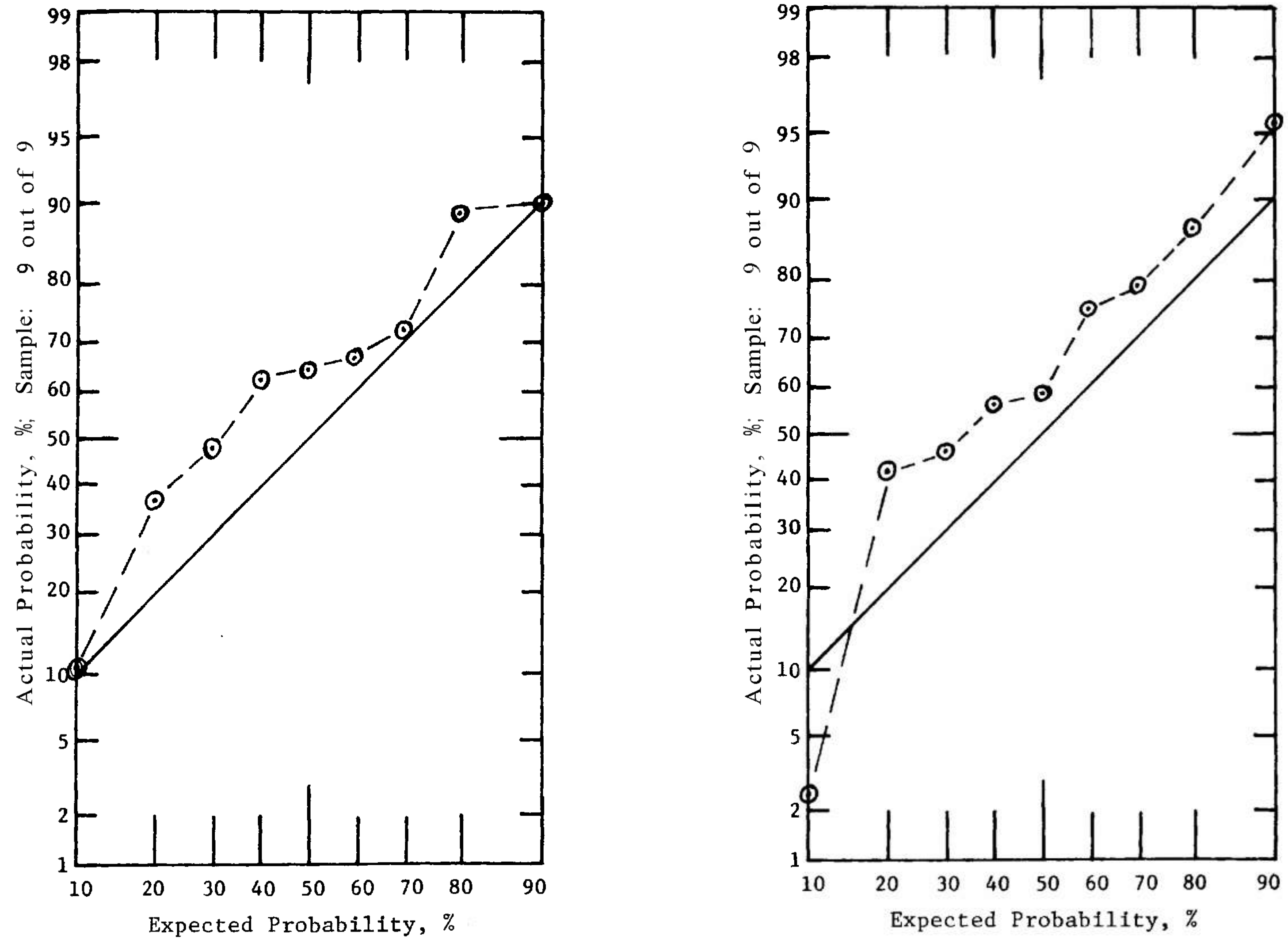


Figure 2-11. (Cont'd)

pening without human assistance are very small. Be suspicious that the data are not fully legitimate.

More information on graphical analysis is given in Refs. 28-29.

2-3 ANALYTIC ESTIMATION OF PARAMETERS OF A DISTRIBUTION

This material is covered in detail in *Part Six, Mathematical Appendix and Glossary.* Some of the material is summarized here for easy reference. It is often worthwhile plotting the sample results from a continuous distribution on graph paper as mentioned in par. 2-2. This ought to be done before analytic estimation.

When the random variable is discrete, s-confidence limits are more difficult to use and their exact statement is more complicated.

2-3.1 BINOMIAL

The discrete random variable k is the number of occurrences of an attribute in N s-independent trials, when the attribute must be in either of 2 mutually exclusive categories.

Use the following notation

k_1, k_2 = number of trials with result #1 or result #2, respectively

p_1, p_2 = probability of result #1 or result #2, respectively, in each and every trial

N = number of s-independent trials

The auxiliary relationships are

$$k_1 + k_2 = N \tag{2-14a}$$

$$p_1 + p_2 = 1 \tag{2-14b}$$

The probability mass function *pmf* is

$$pmf\{k_1, k_2; p_1, p_2, N\} = \frac{N!}{k_1!\,k_2!}\left(p_1^{k_1} p_2^{k_2}\right) \tag{2-15}$$

This symmetric form is easy to remember.

The means and variances are

$$E\{k_i\} = Np_i \text{ for } i = 1, 2 \tag{2-16}$$

$$\mathrm{Var}\{k_i\} = Np_1 p_2 \tag{2-17}$$

The usual statistical problem is to estimate p_1, p_2. Most often it is simplest to estimate the smaller of the two, especially since convenient approximations are applicable when the p_i being estimated is small. In reliability work, the 2 categories are usually Good, Bad (or something equivalent). Since the fraction Bad is ordinarily small (if not, the lack of precision in estimating that fraction is of little concern), it is the parameter to be estimated. The maximum likelihood, unbiased estimate $\hat{p}_i$ of p_i is

$$\hat{p}_i = k_i/N \tag{2-18}$$

s-Confidence limits are somewhat difficult to calculate. If $p_i \leqslant 0.1$ and $N \geqslant 10$, the Poisson approximation (par. 2-3.2) is usually adequate and is easier to use. If this is not feasible, Table 41 of Ref. 30 or Chapter 7 of Ref. 18 will give adequate answers. In that situation ($N \leqslant 10, p \geqslant 0.1$) no engineer really cares exactly what the situation is because (a) probabilities of failure greater than 0.1 are generally bad, and (b) the uncertainty in p is going to be so high that it is usually pointless (for engineers) to make exact calculations. If a contractual relationship is involved, a statistician ought to be consulted. For very rough estimates of uncertainty, estimate $\hat{\sigma}$ from Eq. 2-17 using $p_i = \hat{p}_i$.

$$\hat{\sigma} = N\hat{p}_1\hat{p}_2 \tag{2-19}$$

The s-confidence level associated with a "$\pm 1\sigma$ range" usually will be roughly 50%-80%, (the s-normal tables will not give a correct answer). Example No. 2 illustrates the procedure for obtaining s-confidence limits.

2-3.2 POISSON

The discrete random variable k is the number of events which occur in a fixed set of circumstances. Examples are the number of defects in 3.5 yd^2 of cloth, the number of failures during a 12-hr test, and the number of accidents per driver in 10 yr. If the variable k has the Poisson distribution, the process that generates k is often called a Poisson process. In a Poisson process, it is sometimes convenient to define a rate parameter, e.g., the number of defects per square yard of cloth, the number of failures per hour, or the number of failures per driver per year.

Use the following notation:

*poi** = base-name for Poisson distribution; the * is replaced by *m* to denote the *pmf*, by *f* to denote the *Cdf*, or by *fc* to denote the *Sf*

k = number of events which occur (must be a non-negative integer)

μ = mean number of events

The probability mass function *pmf*, the *Cdf* and *Sf* are

$$poim(k;\mu) = pmf\{k;\mu\} = \exp(-\mu)\mu^k/k! \qquad (2\text{-}20a)$$

$$poif(k;\mu) = Cdf\{k;\mu\} = \sum_{i=0}^{k} poim(i;\mu) \qquad (2\text{-}20b)$$

$$poifc(k;\mu) = Sf\{k;\mu\} = \sum_{i=k}^{\infty} poim(i;\mu) \qquad (2\text{-}20c)$$

The *poif* and *poifc* in *Eq.* 2-20 can be expressed in terms of the chi-square *Cdf* and *Sf*, *csqf* and *csqfc*. (See *Part Six, Mathematical Appendix and Glossary.)*

$$poifc(k;\mu) = csqf(2\mu;2k) \qquad (2\text{-}21a)$$

$$poif(k;\mu) = csqfc(2\mu;2k+2) \qquad (2\text{-}21b)$$

The mean and variance are

$$E\{k\} = \mu \qquad (2\text{-}22a)$$

$$\mathrm{Var}\{k\} = \mu \qquad (2\text{-}22b)$$

The usual statistical problem is to estimate μ. The Poisson distribution often is used as an approximation to the binomial distribution with

$$\mu = pN \qquad (2\text{-}23)$$

and thus the statistical problem may be to estimate p. *Eq.* 2-23 is satisfactory for N large (say, $N \geqslant 10$) and p small (say, $p \leqslant 0.1$), and even works reasonably well when those conditions are violated.

The maximum likelihood, unbiased estimate $\hat{\mu}$ of μ is

$$\hat{\mu} = k \qquad (2\text{-}24)$$

Calculating s-confidence limits for μ is more complicated than when the random variable is continuous. A good explanation is given in *Part Six, Mathematical Appendix and Glossary.* 2-sided s-confidence statements for μ are of the form

$$\mathrm{Conf}\{\mu_L^- \leq \mu \leq \mu_U^+\} \geq C \qquad (2\text{-}25a)$$

Example No. 2

Fifty emergency flares were fired with 46 successes and 4 failures. Estimate the failure probability. Find some s-confidence limits.

Procedure	Example
1. State the experimental data.	1. $N = 50$, $k_1 = 4$, $k_2 = 46$.
2. Estimate $\hat{p}_1$ from Eq. 2-18.	2. $\hat{p}_1 = 4/50 = 8.0\%$.
3. Find s-confidence limits by use of Poisson approximation.	3. Since $N \geq 10$ and $\hat{p} < 0.1$, see par. 2-3.2.
4. From Example No. 3 in par. 2-3.2, calculate the limits for p_1.	4. $p_{1,L-} = 1.37/50 = 2.7\%$; $p_{1,L+} = 1.97/50 = 3.9\%$ $p_{1,U-} = 7.76/50 = 15.5\%$; $p_{1,U+} = 9.16/50 = 18.3\%$.
5. Make the corresponding s-confidence statements.	5. Conf $\{2.7\% \leq p_1 \leq 18.3\%\} \geq 90\%$ Conf $\{3.9\% \leq p_1 \leq 15.5\%\} \leq 90\%$.

One conclusion from this experiment and analysis is that it is difficult to find out much about probabilistic parameters by conducting small experiments.

$$\text{Conf}\{\mu_L^+ \le \mu \le \mu_U^-\} \le C \qquad (2\text{-}25b)$$

where for symmetrical intervals, C' is defined by the equation $1 - C' = (1 - C)/2$, and the s-confidence limits μ_L^-, μ_L^+, μ_U^-, μ_U^+ are defined by

$$poifc(k;\mu_U^-) = csqf(2\mu_U^-; 2k)$$

$$= C', \text{ for } k \neq 0 \qquad (2\text{-}25c)$$

$$1 - poif(k;\mu_U^+) = csqf(2\mu_U^+; 2k + 2) = C' \qquad (2\text{-}25d)$$

and

$$poifc(k;\mu_L^-) = csqf(2\mu_L^-; 2k)$$

$$= 1 - C', \text{ for } k \neq 0 \qquad (2\text{-}25e)$$

$$1 - poif(k;\mu_L^+) = csqf(2\mu_L^+; 2k + 2)$$

$$= 1 - C' \qquad (2\text{-}25f)$$

The case for $k = 0$ is different; symmetrical s-confidence limits have no meaning (because the $poifc$ $(0;\mu) = 1$, regardless of the value of μ. The 1-sided s-confidence statements for μ are of the form

$$\text{Conf}\{\mu \le \mu_U^+\} \ge C \qquad (2\text{-}25g)$$

$$\text{Conf}\{\mu_L^+ \le \mu\} \le C \qquad (2\text{-}25h)$$

where μ_L^+ and μ_U^+ are defined by

$$1 - poif(0;\mu_U^+) = csqf(2\mu_U^+; 2)$$

$$= 1 - \exp(-2\mu_U^+) = C \qquad (2\text{-}25i)$$

$$1 - poif(0;\mu_L^+) = csqf(2\mu_L^+; 2)$$

$$= 1 - \exp(-2\mu_L^+)$$

$$= 1 - C \qquad (2\text{-}25j)$$

The μ^+ and μ^- are seen to be different because, for discrete variables, the *Sf* and Cdf are not complementary. While Eq. 2-25 looks complicated, its application in practice is very straightforward; see Example No. 3.

In most reliability-statistics theory, only **Eq.** 2-25a is given (not Eq. 2-25b) and the inequality often is implied rather than being explicit. It seems wiser to make as much use of the information as we can, and thus to use both the Eqs. 2-25a and 2-25b, and to make the inequalities therein, explicit.

Other kinds of s-confidence statements are feasible although not usually made. For example, in Eqs. 2-25a and 2-25b we could use a single pair of μ_L and μ_U, then use a C^+ in Eq. 2-25a and a C^- in Eq. 2-25b. Example No. 4 illustrates the procedure.

Often it is desired to estimate the Poisson parameter rather than the mean number of events. It is easily found by the following formula

$$\lambda = \mu / t_o \qquad (2\text{-}26)$$

where

λ = Poisson rate parameter (in reliability work, it is virtually always the failure rate)

t_o = fixed characteristic of test, such as length-of-time or area of inspection

Example No. 5 illustrates the procedure.

Now let us consider life test applications. When the failure rate is constant, t_o in Eq. 2-26 is interpreted as the total operating time of all units. it makes no difference what kind of censoring is employed (if any) nor whether the test is with or without replacement; the only requirement is that t_o be fixed in advance of the test. This paragraph applies whether units are tested to

Example No. 3

In a particular test, 4 events were observed. Assume that the process is Poisson. Estimate the true mean and investigate some s-confidence statements for the true mean.

Procedure	Example
1. State the test result. Find a good point estimate of μ.	1. $k = 4$ $\hat{\mu} = k = 4.$
2. Investigate the s-confidence statements; use Eq. 2-25. Choose the s-confidence level. Calculate the degrees-of-freedom for the chi-square tables.	2. Choose $C = 90\%$; so $C' = 95\%$ and $1 - C' = 5\%$. The choice is largely a personal matter. Since $k = 4$, we have $2k = 8$ and $2k + 2 = 10$.
3. Pick the appropriate $\chi^2_{P,\nu}$ values from a table.	3. From Table 2-7, we have $\chi^2_{5\%,8} = 2.733$ for Eq. 2-25e $\chi^2_{5\%,10} = 3.940$ for Eq. 2-25f $\chi^2_{95\%,8} = 15.51$ for Eq. 2-25c $\chi^2_{95\%,10} = 18.31$ for Eq. 2-25d.
4. The μ_L and μ_U are calculated from the appropriate $\chi^2_{P,\nu}$ in step 3.	4. $\mu_L^- = 1.37$, $\mu_L^+ = 1.97$ $\mu_U^- = 7.76$, $\mu_U^+ = 9.16$.
5. Make the two s-confidence statements (from Eqs. 2-25a and 2-25b).	5. $\mathrm{Conf}\{1.37 \leq \mu \leq 9.16\} \geq 90\%$ $\mathrm{Conf}\{1.97 \leq \mu \leq 7.76\} \leq 90\%$.

It is readily seen that the true mean is not known very well.

Example No. 4

In a 1000 hr test, no failures were observed. Assume that the process is Poisson. Estimate the true mean and investigate some s-confidence statements for the mean.

Procedure	Example
1. State the test result.	1. $k = 0$
Find a point estimate of μ.	$\mu = k = 0.$
2. Investigate the s-confidence statements; use Eq. 2-25.	2. Choose C = 95%. Since $k = 0$, the intervals will be 1-sided.
3. Pick appropriate values of μ_L^+ and μ_U^+. Use Eqs. 2-25i and 2-25j. Table 2-7 will suffice.	3. $\mu_L^+ = 0.1026/2 = 0.0513$ $\mu_U^+ = 5.991/2 = 3.00.$
4. Make the two s-confidence statements (from Eqs. 2-25g and 2-25h).	4. $\text{Conf}\{\mu \leq 3.00\} \geq 95\%$ $\text{Conf}\{0.0513 \leq \mu\} \leq 95\%.$

These s-confidence statements in step 4 and the point estimate in step 1 are of little help. Much has been written for this case (no failures) about estimating μ, but much of it is fruitless because it tries to give the illusion of more certainty when there is nothing but vast uncertainty. Randomized s-confidence limits can often considerably narrow the region of uncertainty (see Part *Six Mathematical Appendix and Glossary)*. The two statements in step 4 cannot be combined into one statement with an upper and lower limit; the first one is the one usually given.

Example No. 5

Use data in Example No. 4; estimate the failure rate.

Procedure	Example
1. Calculate the point estimate.	1. $\hat{\lambda} = \hat{\mu}/1000\text{-hr} = 0.$
2. Calculate the interval estimates	2. $\lambda_U^+ = \mu_U^+/1000\text{-hr} = 3.0/1000\text{ hr}$ $\lambda_L^+ = \mu_L^+/1000\text{-hr} = 0.051/1000\text{ hr}.$
3. Make the complete s-confidence statements.	3. $\text{Conf}\{\lambda \leq 3.0/1000\text{ hr}\} \geq 95\%$ $\text{Conf}\{0.051/1000\text{ hr} \leq \lambda\} \leq 95\%.$

TABLE 2-7

THE 5th AND 95th PERCENTILES
OF THE CHI-SQUARE DISTRIBUTION

$csqf\,(\chi^2;\nu) = Cdf\left\{\chi^2;\nu\right\}$

The body of the table gives the values of $\chi^2_{P,\nu}$ such that $csqf\,(\chi^2_{P,\nu};\nu) = P$, for $P = 5\%$ and 95%. The table has been abbreviated (from more extensive tables) for easy use in the examples.

$\nu \backslash P$	5%	95%
2	0.1026	5.991
4	0.7107	9.487
6	1.635	12.59
8	2.733	15.51
10	3.940	18.31
12	5.226	21.03

first failure and then discarded, or whether units are repaired; the only requirement is that λ be a constant, i.e., the process is Poisson. This situation is related to par. 2-3.3 on the exponential distribution since both deal with a Poisson process.

2-3.3 EXPONENTIAL

The continuous random variable t is the time to first failure (time of failure-free operation). It is related to the Poisson process of par. 2-3.2, but instead of counting events (e.g., defects or failures), the time to the first failure (area to first defect) is measured.

Use the following notation:

λ = failure rate (Poisson rate parameter)

t = time to failure

θ = $1/\lambda$ (often used for convenience)

The pdf and Sf are

$$pdf\{t;\lambda\} = \lambda \exp(-\lambda t)$$

$$= (1/\theta)\exp(-t/\theta) \quad (2\text{-}27a)$$

$$Sf\{t;\lambda\} = \exp(-\lambda t) = \exp(-t/\theta) \quad (2\text{-}27b)$$

The mean and variance are

$$E\{t\} = 1/\lambda = \theta \quad (2\text{-}28a)$$

$$\mathrm{Var}\{t\} = 1/\lambda^2 = \theta^2 \quad (2\text{-}28b)$$

The usual statistical problem is to estimate λ (or θ). Tests are often run until k_0 failures have been observed. Then the maximum likelihood, unbiased estimate $\hat\theta$ for θ is

$$\hat\theta = t_k/k_0 \quad (2\text{-}29)$$

where t_k = total operating time up to failure k_0.

The reciprocal of $\hat\theta$ is the maximum likelihood estimator for λ, but it is no longer unbiased.

s-Confidence statements about λ are found from the fact that $2\,k_0\,(\lambda/\hat\lambda)$ has a chi-square distribution with $2\,k_0$ degrees-of-freedom. An equivalent statement is that $\lambda/\hat\lambda$ has a chi-square/nu distribution with $2\,k_0$ degrees-of-freedom. For a 2-sided symmetrical s-confidence interval, we have

$$\mathrm{Conf}\{\lambda_L \le \lambda \le \lambda_U\} = C \quad (2\text{-}30a)$$

where C' is defined by the equation $1 - C' = (1 - C)/2$, and λ_L and λ_U are defined by

$$csqf(2k_0\lambda_U/\hat\lambda;\, 2k_0) = C' \quad (2\text{-}30b)$$

$$csqf(2k_0\lambda_L/\hat\lambda;\, 2k_0) = 1 - C' \quad (2\text{-}30c)$$

1-sided s-confidence intervals are analogous to Eq. 2-30; i.e., ignore λ_U and use **Eq.** 2-30c with C' → C, or ignore λ_L and use Eq. 2-30b with C' → C. Table 2-8 can also be used to find λ_L and λ_U. Example No. 6 illustrates the procedure.

The use of other tables is shown in ***Part Six, Mathematical Appendix and Glossary.*** Several of the tables are more convenient to use. The ratio of $\lambda/\hat{\lambda}$ depends only on k_o, the failure at which the test is stopped. It is very difficult to estimate λ closely when there are few failures; see Table 2-8 for specific information.

In life tests, the total operating time of all units (regardless of how obtained) is used in **Eq. 2-29.** It makes no difference what **kind** of censoring is used (if any) or whether failed items are replaced, or whether failed units are repaired. The only requirements are that λ be a constant, i.e., the process is Poisson, and that the number of failures is not a random variable. This situation is related to par. 2-3.2 on the Poisson process; that paragraph is used when the time is fixed and the number of failures is the random variable. Example No. 7 illustrates the procedure.

Table 2-8 can be used in reverse to find how many failures must be observed for a given "accuracy" in estimation. For example, if a 90% symmetrical s-confidence level is reasonable (95%-5%), then to get a ratio of λ_U/λ_L of about 2 will require about 25-30 failures.

Reliability estimates. One'of the desirable properties of maximum likelihood estimates is that they can be used in any function to give a maximum likelihood estimate of the function. The same is true for s-confidence limits when there is only one parameter being estimated from the sample. Example No. 8 illustrates the procedure.

2-3.4 *s*-NORMAL

The continuous random variable x is the measure of failure resistance, e.g., time-to-failure or stress-to-failure. Occasionally, the physical situation prohibits negative values of the random variable. If the probability of their occurrence is very small, the anomaly is usually ignored; if it can't be ignored, consult a statistician.

TABLE 2-8

***s*-CONFIDENCE LIMITS FOR POISSON RATE PARAMETER (FAILURE RATE)**

For the stated sconfidence, and given that the test was stopped at failure k_0, the body of the table gives the factors for 5% & 95% and for 2.5% & 97.5% Conf $\{\lambda \geq \text{factor} \times \hat{\lambda}\}$ and the ratio of the upper to the symmetrical-lower limit.

k_0	1	2	3	4	5	6	8	10	15	20	50
5% *s*-conf.	3.0	2.4	2.1	1.94	1.83	1.75	1.64	1.57	1.46	1.39	1.24
95% sconf.	0.052	0.18	0.27	0.34	0.39	0.44	0.50	0.54	0.62	0.66	0.78
λ_U/λ_L	58	13	7.7	5.7	4.6	4.0	3.3	2.9	2.4	2.1	1.60
2.5% sconf.	3.7	2.8	2.4	2.2	2.0	1.94	1.80	1.71	1.57	1.48	1.30
97.5% *s*-conf.	0.025	0.12	0.21	0.27	0.32	0.37	0.43	0.48	0.56	0.61	0.74
λ_U/λ_L	146	23	12	8.0	6.3	5.4	4.2	3.6	2.8	2.4	1.75

Example. If the test is run until the second failure, $\lambda \geq 2.4\hat{\lambda}$ in 5% of the experiments and $\lambda \geq 0.18\hat{\lambda}$ in 95% of the experiments; the ratio of the upper to the lower limit is 13. i.e., the true λ is uncertain to within a factor of 13, at the net sconfidence level of 90%.

Example No. 6

Suppose that the total operating time up to the 4th failure was 2024 hr. Estimate λ, θ, and find a suitable s-confidence statement. Assume a Poisson process.

Procedure	Example
1. State the test results.	1. $k_0 = 4$, $t_4 = 2024$ hr.
2. Find a good point estimate for λ, θ. State the degrees-of-freedom ν for the chi-square distribution.	2. $\hat{\theta} = 2024\ \text{hr}/4 = 506$ hr. $\hat{\lambda} = 1/\hat{\theta} = 1/506\text{-hr} = 1.98/1000\text{-hr}$. $\nu = 2k_0 = 8$.
3. Choose a s-confidence level.	3. $C = 90\%$ is reasonable; the choice is very subjective. $C' = 95\%$, $1 - C' = 5\%$.
4. Find the values of $\chi^2_{P,\nu}$.	4. Use Table 2-7. $\chi^2_{5\%,8} = 2.733$, for Eq. 2-30c $\chi^2_{95\%,8} = 15.51$, for Eq. 2-30b.
5. Find the corresponding λ_L and λ_U, the lower and upper s-confidence limits for λ.	5. $\lambda_L = [\chi^2_{5\%,8}/(2k_0)]\hat{\lambda} = 0.68/1000\text{-hr}$ $\lambda_U = (15.51/8)(1.98/1000\text{-hr}) = 3.8/1000\text{-hr}$.
6. Make the s-confidence statement.	6. $\text{Conf}\{0.68/1000\text{-hr} \le \lambda \le 3.8/1000\text{-hr}\} = 90\%$.
7. Use Table 2-8 to check the results.	7. $\lambda_L = 0.34\hat{\lambda} = 0.67/1000\text{-hr}$ $\lambda_U = 1.94\hat{\lambda} = 3.8/1000\text{-hr}$ $\lambda_U/\lambda_L = 5.7$. The results are within rounding errors of step 5.

Example No. 7

A helicopter inertial navigation system is tested (Ref. 11) and truncated at 4 failures, at which time 1500 hr of test time have elapsed. Estimate the mean life, the 2-sided (symmetrical) upper and lower s-confidence limits, and the 1-sided lower s-confidence limit on mean life for a s-confidence level of 0.95.

Procedure	Example
1. Tabulate the test results.	1. t = time elapsed at the k_0-th failure = 1500 hr k_0 = number of failures = 4.
2. Compute the estimated mean life from $\hat{\theta} = t/k_0$ (2-29)	2. $\hat{\theta}$ = 1500 hr/4 = 375 hr $\hat{\lambda} = 1/\hat{\theta}$ = 2.67/1000-hr.
3. Calculate the symmetrical levels.	3. C = 95% C' = 97.5%, $1 - C'$ = 2.5%.
4. Copy the appropriate data from Table 2-8.	4. Ratio $\lambda/\hat{\lambda}$ — s-confidence 2.2 — 2.5% 0.27 — 97.5% 1.94 — 5%
5. Compute s-confidence limits for 95% symmetric levels.	5. $\lambda_U = 2.2\hat{\lambda} = 2.2 \times 2.67$/1000-hr = 5.87/1000-hr $\theta_L = 1/\lambda_U$ = 170 hr $\lambda_L = 0.27\hat{\lambda} = 0.27 \times 2.67$/1000-hr = 0.72/1000-hr $\theta_U = 1/\lambda_L$ = 1390 hr.
6. Compute the 1-sided 95% s-confidence limit.	6. $\lambda_U = 1.94\hat{\lambda} = 1.94 \times 2.67$/1000-hr = 5.18/1000-hr $\theta_L = 1/\lambda_U$ = 193 hr.
7. Make the s-confidence statements.	7. Conf{170 hr $\leq \theta \leq$ 1390 hr} = 95% Conf{193 hr $\leq \theta$} = 95%.

Example No. 8

Same data as Example No. 7, but find the corresponding estimates for *s*-reliability with a mission time of 50 hr.

Procedure	Example
1. State $\hat{\lambda}$ and s-confidence limits for λ.	1. $\hat{\lambda} = 2.67/1000\text{-hr}$ $\text{Conf}\{0.72/1000\text{-hr} \le \lambda \le 5.87/1000\text{-hr}\} = 95\%$ $\text{Conf}\{\lambda \le 5.18/1000\text{-hr}\} = 95\%.$
2. Calculate $\hat{R}$. (Eq. 2-27b)	2. $\hat{R} = \exp(-\hat{\lambda}t)$ $= \exp[-(2.67/1000\text{-hr}) \times 50 \text{ hr}]$ $= 0.875$
3. Calculate 95% (symmetrical) s-confidence limits for R.	3. $R_U = \exp(-\lambda_L t)$ $= \exp[-(0.72/1000\text{-hr}) \times 50 \text{ hr}]$ $= 0.965$ $R_L = \exp(-\lambda_U t)$ $= \exp[-(5.87/1000\text{-hr}) \times 50 \text{ hr}]$ $= 0.746$ $\text{Conf}\{0.746 \le R \le 0.965\} = 95\%.$
4. Calculate the 95% lower 1-sided limit for R.	4. $R = \exp(-\lambda_U t)$ $= \exp[-(5.18/1000\text{-hr}) \times 50 \text{ hr}]$ $= 0.772$ $\text{Conf}\{R \ge 0.772\} = 95\%.$

Use the following notation:

μ = mean (location parameter)

σ = standard deviation (scale parameter)

x = random variable

N = sample size

x_i = value of x for item i in sample

The *pdf* and *Cdf* are

$$pdf\{x;\mu,\sigma\} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{2-31a}$$

$$Cdf\{x;\mu,\sigma\} = gauf\left(\frac{x-\mu}{\sigma}\right) \tag{2-31b}$$

The usual statistical problem is to estimate μ and σ from a sample.

It is convenient to classify experiments according to whether or not the complete sample was failed.

2-3.4.1 All Items Teated to Failure

The maximum likelihood, unbiased estimate $\hat{\mu}$ of μ is

$$\hat{\mu} = \text{sample mean} = \bar{x} \equiv (1/N) \sum_{i=1}^{N} x_i \tag{2-32}$$

The sample median is also a very good estimate of μ.

The maximum likelihood estimate for the standard deviation is the sample standard deviation.

$$\hat{\sigma} = \left[\frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2\right]^{-1/2} \tag{2-33}$$

$\hat{\sigma}^2$ is the maximum likelihood estimate for σ^2. The unbiased estimate s^2 for the variance σ^2 is

$$s^2 = \hat{\sigma}^2 \times \left(\frac{N}{N-1}\right) \tag{2-34}$$

The s-statistic in **Eq.** 2-34 is also very useful in many statistical tests involving the s-normal distribution; s is NOT an unbiased estimator for σ. The good reason s is used so much, rather than $\hat{\sigma}$, is that the sampling distribution of s is known well, whereas that of $\hat{\sigma}$ is not.

The *Cdf* of $\bar{x}$ is

$$Cdf\{\bar{x};\mu,\sigma\} = gauf\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{N}}\right) \tag{2-35}$$

The *Cdf* of s is

$$Cdf\{s;\mu,\sigma\} = csqf\left[\frac{(N-1)s^2}{\sigma^2};N-1\right] \tag{2-36}$$

Eq. *2-35* is not feasible to use because σ is not known. The Student *t*-distribution is used instead (see Ref. 30).

$t = \frac{\bar{x}-\mu}{s/\sqrt{N}}$ has the Student t-distribution with $N-1$ degrees of freedom (2-37)

Example No. 9 illustrates the application of the Student t-distribution.

2-3.4.2 Censored Samples

The simple approach in par. **2-3.4.1** cannot be used for censored samples. Maximum likelihood methods are most usual for estimation in this case and can give an idea of the uncertainties involved. Part **Six** or Ref. 13 or a programmer/statistician ought to be

Example No. 9

For Data Set A (Table 2–3) find the sample mean and sample standard deviation. Estimate μ and σ, and find suitable s-confidence limits for each.

Procedure	Example
1. Calculate the sample mean by Eq. 2–32. Estimate $\hat{\mu}$.	1. $\bar{x} = 1950.7$ hr $\hat{\mu} = 1951$ hr.
2. Calculate the sample standard deviation by Eq. 2–33.	2. $\hat{\sigma} = 859.0$ hr.
3. Calculate s by Eq. 2–34.	3. $s = \hat{\sigma}\sqrt{\frac{N}{N-1}} = 859.0 \text{ hr} \sqrt{\frac{20}{19}}$ $= 881.3$ hr.
4. Find the 90%, 2–sided (symmetric) s-confidence limits for μ. Use Eq. 2–37 and the t tables in ***Part Six*** or in Ref. 30 or elsewhere. μ_U, μ_L are the upper and lower limits for μ.	4. $t = \pm 1.729$ for 95%, and 5% and 19 degrees of freedom. $\pm 1.729 = \frac{\bar{x} - \mu}{s/\sqrt{N}} = \frac{1950.7 \text{ hr} - \mu}{881.3/\sqrt{20}}$ $\mu_U = \left(\frac{881.3}{\sqrt{20}} \times 1.729 + 1950.7\right)$ hr $= (340.7 + 1950.7)$hr ≈ 2290 hr $\mu_L = \left[\frac{881.3}{\sqrt{20}} \pm (-1.729) + 1950.7\right]$hr $= (-340.7 + 1950.7)$hr ≈ 1610 hr.
5. Compare with the result of the graphical analysis in par. 2–2.3 Example No. 1(A) step 3. They are reasonably close.	5. Graphical estimation: $1400 \text{ hr} \leq \mu \leq 2440$ hr, $\hat{\mu} = 1920$ hr. Analytic estimation: $\text{Conf}\{1610 \text{ hr} \leq \mu \leq 2290 \text{ hr}\} = 90\%$ $\hat{\mu} = 1951$ hr.

Example No. 9 (Contfd)

6. Find the 90% 2-sided (symmetric) s-confidence limits for σ. Use Eq. 2-36 and chi-square tables in ***Part Six*** or Ref. 30, or elsewhere.	6. $csqf(10.1; 19) = 5\%$ $csqf(30.1; 19) = 95\%$ (from the tables) $\frac{(N-1)s^2}{\sigma_L^2} = 30.1 = \frac{19 \pm (881.3 \text{ hr})^2}{\sigma_L^2}$ $\sigma_L = 700.2$ hr $\frac{(N-1)s^2}{\sigma_U^2} = 10.1 = \frac{19 \times (881.3 \text{ hr})^2}{\sigma_U^2}$ $\sigma_U = 1208.0$ hr.
7. Compare with the result of the graphical analysis in par. 2-2.3, Example No. 1(A) step 4. The point estimates are close enough.	7. Graphical estimation: $290 \text{ hr} \leq \sigma \leq 2960 \text{ hr}$ $\hat{\sigma} = 810$ hr. Analytic estimation: $\text{Conf}\{700 \leq \sigma \leq 1210 \text{ hr}\} = 90\%$ $\hat{\sigma} \simeq 860$ hr.

consulted. Many statistical packages for computers have prepared programs for making this calculation.

2-3.4.3 s-Confidence Limits for s-Reliability

Because two parameters (instead of one) have been estimated from the data, s-confidence limits are virtually impossible to calculate for the s-reliability. A statistician ought to be consulted in **this** case. Perhaps Prediction Intervals or Tolerance Intervals can be used.

2-3.5 WEIBULL

The continuous random variable t is the measure of failure resistance, e.g., time-to-failure or stress-to-failure.

Use the following notation:

t = random variable (e.g., time to failure)

α = scale parameter (characteristic value)

β = shape parameter

The Sf is

$$weifc(t/x;\beta) \equiv Sf\{t;\alpha,\beta\} = \exp\left[-(t/\alpha)^{\beta}\right] \quad (2\text{-}38)$$

The mean and variance are

$$E\{t\} = \alpha\Gamma(1 + 1/\beta) \quad (2\text{-}39a)$$

$$\mathrm{Var}\{t\} = \alpha^2\{\Gamma(1 + 2/\beta) - [\Gamma(1 + 1/\beta)]^2\} \quad (2\text{-}39b)$$

The usual statistical problem is to estimate a and β from a sample. Unfortunately, there are no good, simple estimators for a, β. See Ref. 15 or ***Part Six*** or a statistician/programmer. Maximum likelihood methods often are used because they allow most any kind of censoring, and provide a measure of the uncertainties and correlations in the estimates of a and β.

2-3.6 LOGNORMAL

The continuous random variable t is the measure of failure resistance, e.g., time to failure or stress-to-failure.

Use the following notation:

t = random variable (e.g., time to failure)

α = scale parameter (median)

β = shape parameter

μ = mean of $\ln t$

σ = standard deviation of $\ln t$

$B = \exp\left[1/(2\beta^2)\right]$

The Cdf is

$$Cdf\{t;\alpha,\beta\} = gauf[\ln (t/\alpha)^{\beta}] \quad (2\text{-}40)$$

The mean and variance are

$$E\{t\} = \alpha B \quad (2\text{-}41a)$$

$$\mathrm{Var}\{t\} = \alpha^2 B^2 (B^2 - 1) \quad (2\text{-}41b)$$

The usual statistical problem is to estimate a and β from a sample. Probably the simplest procedure is to take natural logs of all the data and to proceed as if the distribution were s-normal. The unbiased property of the estimators will disappear, but the maximum likelihood property and s-confidence limit transferance remain.

$$\alpha = \exp(\mu) \quad (2\text{-}42a)$$

$$\beta = 1/a \quad (2\text{-}42b)$$

See par. 2-3.4 for the s-normal distribution.

2-4 GOODNESSOF-FIT TESTS

A most important consideration is why the test is being performed; see par. 2-1 for a full discussion of this point. The two goodness-of-fit tests described in this paragraph make a null hypothesis, i.e., the sample is from the assumed distribution. Then a statistic, evaluated from the sample data, is calculated and looked-up in a table that shows how lucky/unlucky you were for that sample. The luck is determined by the size **of** the 2-sided tail area. If that tail area is very small (you were very unlucky if the null hypothesis is true), the null hypothesis (there is no difference between the actual and the assumed distributions) is rejected. Otherwise, the null hypothesis is accepted, i.e., the actual distribution could easily have generated that set of data (within the range of the data); the test says nothing about the behavior of the distribution outside the range of the data.

There are many goodness-of-fit tests (Refs. **3,** 4, 6, 7, 24, and 25). The two presented in this paragraph are all-purpose (do not depend very much on which distribution is assumed) and are reasonably good. The chi-square test ought not be used for too-small samples (say, less than 30) because the assumptions involved are not likely to be fulfilled.

Ref. **24** (Chap. **30)** and Ref. 25 discuss both tests. In practice, many of the important requirements are not fulfilled because too much of the analysis is decided after seeing the data. But if the test is used for "ballpark" confirmation, little harm is done.

Tests-of-fit are statistical tests, not engineering tests. No matter what the distribution or what the test, it is possible to take a sample small enough ***so*** that virtually no distribution will be rejected, or large enough so that virtually every distribution will be rejected.

Tests-of-fit do NOT determine how well the proposed distribution will fit the actual one in the regions where there are no data. It is poor practice to find the "best" distribution by choosing the one which fits the sample best. The examples in Figs. 2-10 and 2-1 1 ought to dispell that notion once and for all.

2-4.1 CHI-SQUARE TEST

The chi-square test is performed by dividing the data from the sample into cells. The actual number of data points in each cell is compared to the predicted number for that cell and a combined statistic X^2 is calculated for all cells; it is then compared with χ_ν^2.

$$X^2 \equiv \sum_{i=1}^{k} \frac{(n_0 - n_e)^2}{n_e} \tag{2-43}$$

where

χ_ν^2 = a random variable having the chi-square distribution with ν degrees of freedom

n_0 = observed number in each cell

n_e = expected number in each cell

k = number of cells

ν = $k - 1$ if none of the parameters of the distribution is estimated from the data. If s parameters are estimated from the data, then $k - 1 \geqslant \nu \geqslant k - 1 - s$. The exact value of ν depends on how the parameter estimates were made.

If n_e is the same for each cell (equal probability method of choosing cells), then $n_e = N/k$ and

$$X^2 = \left(\frac{k}{N} \sum_{i=1}^{k} n_0^2\right) - N \tag{2-44}$$

where N = number in sample.

A heuristic (and rigorously erroneous, but very useful) description of the source of the statistic is helpful (see Ref. **24,** Chapter 30 for detailed derivations).

The number in each cell has a binomial distribution, which depends on 2 parameters: the total number in the sample, and the probability of a value falling in that cell. As the s-expected number in the cell becomes very large, the binomial distribution turns into a s-normal distribution with the same mean n_e and a standard deviation equal to the square root of the mean $\sqrt{n_e}$. The number in each cell is converted to a standard s-normal variate by subtracting the mean and dividing by the standard deviation. The sum of the squares of such variates has a chi-square distribution with $k - 1$ degrees of freedom; each term is of the form

$$\left(\frac{n_0 - n_e}{\sqrt{n_e}}\right)^2 \tag{2-45}$$

The 1 degree of freedom is lost because the last variate is not s-independent, i.e., it can be calculated from the previous data because the total number in the sample is **known.**

Conventionally, only a I-tail test is used, i.e., the calculated value of X^2 ought not be too large. But one ought to be equally suspicious of too-small values; if the usual variability is not there, someone may have tampered with the data. Conventional wisdom suggests that there ought to be at least **5** data points in each cell and at least 30 total data points. But the usual engineer will do the best he can with the data he has.

If the calculated value of X^2 is greater than the tabulated value for χ_ν^2 (at a particular s-significance level) reject the assumed distribution.

A table of $\chi^2_{p,\nu}$ values is given in Table 2-9; other tables are given in Ref. 30.

Example No. 10 illustrates the application of the Weibull distribution.

2-4.2 THE KOLMOGOROV-SMIRNOV (K-S) TEST

The Kolmogorov-Smirnov test is another analytic procedure for testing goodness-of-fit. The procedure compares the observed distribution with a completely-specified hypothesized-distribution and finds the maximum deviation between the *Cdfs* for the two. This deviation is then compared with a critical value that depends on a preselected level of s-confidence (Refs. 7, 19, **24,** and 25); see Table 2-12.

The Kolmogorov-Smirnov test is distribution-free; it can be used regardless of the failure distribution that the data are assumed to follow, provided the random variable is continuous. The discriminating ability of the test depends on the sample size; larger sizes discriminate better. If the random variable is discrete, the s-confidence level will be greater than that shown in Table **2-12** (Ref. **24).**

The test is good regardless of sample size. Most discussions of the discriminating ability of the test are from a statistical rather than an engineering viewpoint.

The steps in a Kolmogorov-Smirnov test are as follows.

1. Completely specify the hypothetical distribution to be tested, $F_{hyp}(x)$. If it has several parameters, a value for each of those parameters must be specified. If any of the parameters were estimated from the data, step **4** must be modified as specified in step **4.**

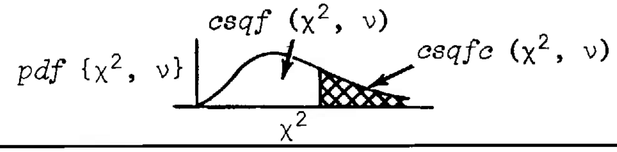
2. At each sample point x_i:

[text continues on page 2-69]

TABLE 2-9

COMPLEMENT OF *Cdf* OF χ^2 (Adapted from Ref. 3)

Body of **the** table gives the values of χ^2.



ν	csqfc (χ^2, ν) 0.99	0.975	0.95	0.90	0.80	0.75	0.50	0.25	0.20	0.10	0.05	0.025	0.01	0.001
1	,000157	,000982	0.00393	0.0158	0.0642	0.10153	0.455	1.323	1.642	2.706	3.841	5.024	6.635	io.827
2	0.0201	0.0506	0.103	0.211	0.446	0.5753	1.386	2.772	3.219	4.605	5.991	7.377	9.210	13.815
3	0.115	0.216	0.352	0.584	1.005	1.2125	2.366	4.108	4.642	6.251	7.815	9.348	11.345	16.268
4	0.297	0.484	0.711	1.064	1.649	1.9225	3.357	5.385	5.989	7.779	9.488	11.143	13.277	18.465
5	0.554	0.831	1.145	1.610	2.343	2.674	4.351	6.625	7.289	9.236	11.070	12.832	15.086	20.517
6	0.872	1.237	1.635	2.204	3.070	3.454	5.348	7.840	8.558	10.645	12.592	14.449	16.812	22.457
7	1.239	1.689	2.167	2.833	).822	4.254	6.346	9.037	9.803	12.017	14.067	16.013	18.475	24.322
8	1.646	2.179	2.733	3.490	4.594	5.070	7.344	10.218	11.030	13.362	15.507	17.534	20.090	[illegible]
9	2.083	2.700	3.325	4.168	5.380	5.898	8.343	11.388	12.242	14.684	16.919	19.023	21.666	29.588
10	2.558	3.247	3.940	4.865	6.179	6.737	9.342	12.548	13.442	15.987	18.307	20.483	23.209	29.5
11	3.053	3.816	4.575	5.578	6.989	7.584	10.341	13.701	14.631	17.275	19.675	21.920	24.725	31.264
12	3.571	4.404	5.226	6.304	7.807	8.438	11.340	14.845	15.812	18.549	21.026	23.336	26.217	32.909
13	4.107	5.008	5.892	7.042	8.634	9.299	12.340	15.984	16.985	19.812	22.362	24.735	27.680	34.528
14	4.660	5.628	6.571	7.790	9.467	io.165	13.339	17.117	18.151	21.064	23.685	[illegible]	29.141	36.123
15	5.229	6.262	7.261	8.547	10.307	11.036	14.339	18.245	19.311	22.307	24.996		30.578	37.697
16	5.812	6.907	7.962	9.312	11.152	11.912	15.338	19.368	20.465	23.542	26.296	28.845	32.000	39.252
17	6.408	7.564	8.672	L0.085	12.002	12.791	16.338	20.488	21.615	24.769	27.587	30.191	33.409	40.790
18	7.015	8.231	9.390	to. 865	12.857	13.675	17.338	21.605	22.760	25.989	28.869	31.526	34.805	42.312
19	7.633	8.906	.0.117	11.651	13.716	14.562	18.338	22.717	23.900	27.204	30.144	32.852	36.191	43.820
20	8.260	9.591	.0.851	12.443	14.578	15.452	19.337	23.827	25.038	28.412	31.410	34.169	37.566	45.315
21	8.897	.0.283	.1.591	13.240	15.445	16.344	20.337	24.935	26.171	29.615	32.671	35.479	38.932	46.797
22	9.542	.0,982	.2.338	14.041	16.314	17.239	21.337	26.039	27.301	30.813	33.924	36.780	40.289	48.268
23	.0.196	.1.688	.3.091	14.848	17.187	18.137	22.337	27.141	28.429	32.007	35.172	38.075	41.638	49.728
24	.0.856	.2.400	.3.848	15.659	18.062	19.037	23.337	28.241	29.553	33.196	36.415	39.364	42.980	51.179
25	.1.524	.3.119	.4.611	.6.473	18.940	19.939	24.337	29.339	30.675	34.382	37.652	40.646	44.314	52.623
26	2.198	.3.844	.5.379	.7.292	19.820	20.843	25.336	30.434	31.795	35.563	38.885	[illegible]923	45.642	54.052
27	.2.879	.4.573	6.151	.8.114	20.703	21.749	26.336	31.528	32.912	36.741	40.113	43.194	46.963	55.476
28	3.565	.5.308	6.928	.8.933	21.588	22.657	27.336	32.620	34.027	37.916	41.337	44.460	48.278	56.893
29	4.256	.6.047	7.708	.9.768	22.475	23.566	28.336	33.711	35.139	39.087	42.557	45.722	49.588	58.302
30	4.953	.6.791	8.493	20.599	23.364	24.476	29.336	34.799	36.250	40.256	43.773	46.980	50.892	59.703

For ν > 30, the quantity $\sqrt{2\chi^2}$ is approximately s-normally distributed with mean $\sqrt{2\nu - 1}$ and variance 1.

Example No. 10

A group of 50 relays is life tested (Ref. 4). The numbers of cycles to failure are given in Table 2-10.

TABLE 2-10

CYCLES TO FAILURE

00	10	20	30	40
1283	4865	8185	13167	28946
1887	5147	8559	14833	29254
1888	5350	8843	14840	30822
2357	5353	9305	14988	38319
3437	5410	9460	16306	41554
3606	5536	9595	17621	42870
3752	6499	10247	17807	62690
3914	6820	11492	20747	63910
4394	7733	12937	21990	68888
4398	8025	12956	23449	73473

Because it often is assumed that relay life data have a Weibull distribution, we will estimate the parameters of the Weibull distribution which fit the data; then we will test the hypothesis that the data came from that exact distribution.

Procedure	Example
1. Estimate the parameters by a graphical method (details not given here) c = cycles-to-failure. (Eq. 2-38)	1. $Sf\{c\} = weifc\,(c;\alpha,\beta) = \exp[-(c/\alpha)^{\beta}]$ $\hat{\beta} = 1.2$ $\hat{\alpha} = 16.6 \times 10^3$ cycles $N = 50$.
2. State the number of points. Choose the number of cells k. There are 50 points, and it is nice to have an s-expected number of points in each cell of 5 or more.	2. $k = 50/5 = 10$.
3. Calculate the cell boundaries. Use the "equal probability" method because it is handy, and does not depend on the data. The range of 0-1 for Sf is divided into k equal parts.	3. The Sf cell boundaries are $i \times \frac{1}{10}$, $i = 0, 1, 2, \ldots, 10$: 0.0, 0.1, 0.2, 0.3, ..., 0.9, 1.0.

Example No. 10 (Cont'd)

4. Calculate the values of c which form the boundaries

$c = 16.6k\left(-\frac{\ln Sf}{1.2}\right)$ (the inverse of **Eq.** 2-38).

From the data find how many are in each cell. The n_e is 5, because the $(\Delta Sf) \times 50 = 5$.

4. See Table 2-11 for results.

5. Find X^2, ν. Use **Eq.** 2-44 for X^2 $k - 1 \geq \nu \geq k - s - 1$. 2 parameters ($\hat{\alpha}$ and $\hat{\beta}$) were determined.

5. $X^2 = 9.20$ for the sample (same answer as in Table 2-11).

$s = 2$, $k = 10$

$9 \geqslant \nu \geqslant 7$.

6. Find $csqf(X^2, \nu)$ from Table 2-11. Numbers very near one cause the null hypothesis to be rejected.

6. $csqf(9.20; 0) \approx 0.58$

$csqf(9.20; 7) \approx 0.75$

Thus, values of X^2 would, by chance alone, be bigger than 9.2 about 25% to 40% of the time.

7. Is the result s-significant; i.e., how unlucky were we?

7. Our value of 9.2 is reasonable and we do not reject the null hypothesis that "the lives in Table 2-10 are from the Weibull distribution calculated in step 1." The results are not s-significant even at the 20% level.

TABLE 2-11

CALCULATIONS FOR RELAY FAILURE PROBLEM

Cell Boundaries		Number in Cell		
Sf	C (k-cycles)	Observed n_o	Expected n_e	$\frac{(n_o - n_e)^2}{n_e}$
1.0	0.0	1	5	3.20
0.9	1.46	3	5	0.80
0.8	3.09	7	5	0.80
0.7	4.93	7	5	0.80
0.6	7.07	7	5	0.80
0.5	9.59	3	5	0.80
0.4	12.68	7	5	0.80
0.3	16.65	4	5	0.20
0.2	22.26	4	5	0.20
0.1	31.85	7	5	0.80
0	∞	–	–	–
		50	50	9.20

a. Compute the hypothetical Cdf as $F_{hyp}(x_r)$.

b. Compute the sample Cdf as $F_{Hi} \equiv r/N$ and as $F_{Lo} \equiv (r-1)/N$, where N is the sample size, and r is the order number of the sample point.

c. Calculate d_r, the absolute value of the maximum difference between the sample and hypothetical Cdf's. Do this by finding $|F_{hyp}(x_r) - F_{Hi}|$ and $|F_{hyp}(x_r) - F_{Lo}|$; the larger one is d_r.

3. Find $d_{r,max}$, the largest d_r (over all r) in step 2c.

4. Find the critical value d from Table 2-12 for the sample size and the selected s-confidence level. If $d_{r,max} \leqslant d$, accept the hypothesis that the observed sample could have come from the hypothetical distribution $F_{hyp}(x)$; otherwise, reject the hypothesis. If the s-confidence level is C, the correct decision will have been made on the fraction C of the occasions the test is used when the hypothesis is true. If the hypothesis is not true, it is complicated to find the fraction of occasions the correct decision will be made.

In many cases, the parameters of the hypothetical Cdf will be estimated from the sample (test) data. Under these circumstances the critical d-values in Table 2-12 are too large and will lead to higher s-confidence levels than anticipated (higher than specified in the table). Results of Monte Carlo investigations have shown that the following rule-of-thumb adjustments to Table 2-12 can be made to yield good critical values for the s-normal and exponential distributions (Ref. 3).

5. In step 4 when estimating mean and standard deviation of a s-normal distribution from the test data, multiply the value of d from Table 2-12 by 0.67. When estimating the mean life for an exponential distribution, multiply the value of d from Table 2-12 by 0.80.

Another way of using the Kolmogorov-

TABLE 2-12

CRITICAL VALUES d OF THE MAXIMUM ABSOLUTE DIFFERENCE BETWEEN SAMPLE AND POPULATION FUNCTIONS FOR THE 2-SIDED K-S TEST (ADAPTED FROM REF. 3)

Sample Size N	s-Confidence Level 00%	05%	90%	95%	99%
4	0.49	0.52	0.56	0.62	0.73
5	0.45	0.47	0.51	0.56	0.67
10	0.32	0.34	0.37	0.41	0.49
15	0.27	0.28	0.30	0.34	0.40
20	0.23	0.25	0.26	0.29	0.36
25	0.21	0.22	0.24	0.27	0.32
30	0.19	0.20	0.22	0.24	0.29
35	0.18	0.19	0.20	0.23	0.27
40	0.17	0.18	0.19	0.21	0.25
50	0.15	0.16	0.17	0.19	0.23
$N \geqslant 10$	$\frac{1.07}{\sqrt{N+1}}$	$\frac{1.14}{\sqrt{N+1}}$	$\frac{1.22}{\sqrt{N+1}}$	$\frac{1.36}{\sqrt{N+1}}$	$\frac{1.63}{\sqrt{N+1}}$

Smimov test is to find the critical value of d from Table 2-12 first. Then add it to and subtract it from the sample *Cdf* (the F_{Hi} and F_{Lo} in step 2b). This gives a band within which the hypothetical distribution will lie (at the stated s-confidence level). This is a very good approach, especially when the sample **is** plotted on special graph paper for which the hypothetical distribution will be a straight line. It is explained more fully in par. 2-5.

The procedure is illustrated in Fig. 2-12(A) where the sample data from Table **2-13** are plotted. The line **S** is the result of plotting F_{Hi} and F_{Lo} from step 2b. Lines **U** and **L** are obtained by adding *d* to and subtracting *d* from line **S**, respectively; *d* = **29%** from Table **2-1**2 for N = 20 and s-confidence = **95%.** The true distribution (uniform over zero to one) **is** the 45-deg line from the origin to the point (1,100%); it lies well within the s-confidence band bounded by the lines **U** and **L.**

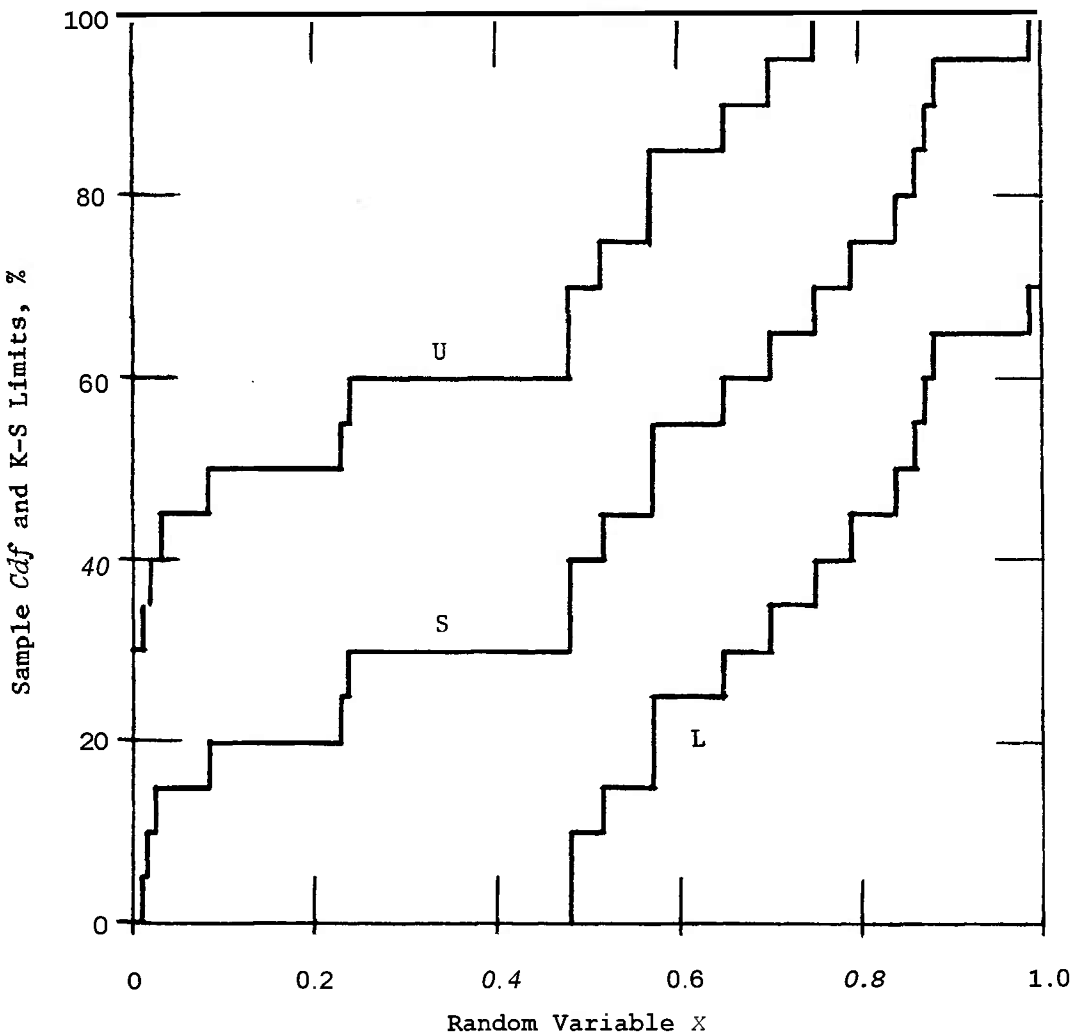


Figure 2-12 (A). Kolmogorov-Smirnov Limits (95% s-Confidence and Sample Cdf—from Table 2-13)

TABLE 2-13

RANDOM SAMPLE FROM THE UNIFORM DISTRIBUTION

These numbers were taken from a table of pseudo-random numbers which were uniformly distributed between 0.00 and 1.00. They have been ordered from smallest to largest. N (sample size) = 20; r is the order number; x is the random variable.

r	x	r	x	r	x	r	x
1	0.01	6	0.24	11	0.57	16	0.84
2	0.02	7	0.48	12	0.65	17	0.86
3	0.03	8	0.48	13	0.70	18	0.87
4	0.08	9	0.52	14	0.75	19	0.88
5	0.23	10	0.57	15	0.79	20	0.99

The sample Cdf and the 95% s-confidence limits (see Table 2-12) are plotted in Fig. 2-12(A). Each point in the table is plotted at $(r-1)/N$ and at r/N for the sample *Cdf.*

It is easy to see that the sample could easily have come from many other distributions, i.e., any that lie between lines U and L.

In practice, plotting the graphs will be inaccurate and/or tedious for $N > 10$, say. Therefore, the shortcut analytic method (described in the next paragraph) ought to be used. It is based on the fact that, even at poor (low) levels of s-confidence the critical value of d is large compared to $1/N$ ($1/N$ is the increase in the sample *Cdf* at each sample point). At each "evaluated sample-point" it is reasonably easy to calculate the smallest subsequent sample-point which could possibly cause rejection of the hypothetical distribution.

Fig. 2-12(B) shows how this exercise is done. Suppose the sample-point number $i = i_0$ has been evaluated and is within the lines U and L as described in Fig. 2-12(A). Fig. 2-12(B) is a small portion of a typical sample *Cdf* plot with the K-S lines U and L shown on it. $F_{hyp}(x)$ can be rejected only if it crosses lines U or L; we will find the smallest sample-point for which that can happen.

If $F_{hyp}(x)$ is to be rejected by crossing line L (see path **A**), the earliest it can do *so* is for the smallest sample-point number i_A for which

$$\left(i_A/N\right) - d \geq F_{hyp}\left(x_{i_0}\right) \tag{2-46}$$

This is so because a *Cdf* can never decrease.

If $F_{hyp}(x)$ is to be rejected by crossing line U (see path **B**), the earliest it can do *so* is for the smallest sample-point number i_B for which

$$F_{hyp}\left(x_{i_B}\right) \geq \left(i_0/N\right) + d \tag{2-47}$$

This is so because the sample *Cdf* increases by $1/N$ at each sample point.

The next sample-point number is the smaller of i_A and i_B. Example No. 11 illustrates the procedure.

2-5 KOLMOGOROV-SMIRNOV *s*-CONFIDENCE LIMITS

Chapter 30 of Ref. 24 (and other references) shows that the **K-S** critical values in

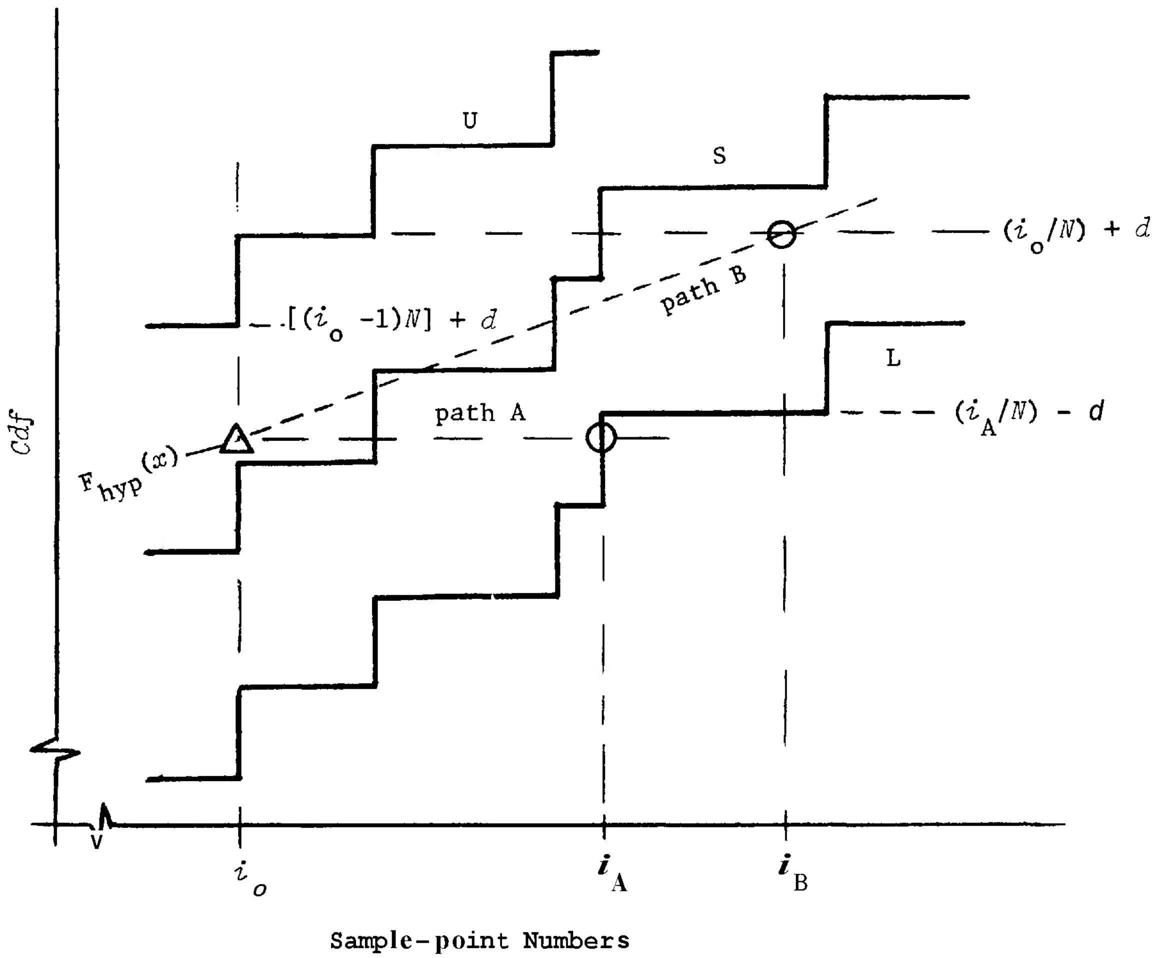


Figure 2-12(B). Kolmogorov-Smirnov Goodness-of-Fit Jest, Shortcut Calculation

Table 2-12 can be used to put s-confidence limits on the actual *Cdf.* The steps in the procedure are:

1. Select the desired s-confidence level and pick the value of d from Table 2-12. It will also depend on the sample size N. A s-confidence level of about $1 - (1/N)$ is reasonable.

2. At each sample-point number r:

a. Compute the sample *Cdf* as:

$$F_{Hi} = r/N, \quad F_{Lo} = (r - 1)/N \qquad (2\text{-}48)$$

This is a pair of points on line *S;* see Fig. 2-12.

b. Compute $F_{Hi} + d$, and $F_{Lo} + d$. This is a pair of points on line U, the upper s-confidence line; see **Fig.** 2-12.

c. Compute $F_{Hi} - d$, and $F_{Lo} - d$. This is a pair of points on line L, the lower s-confidence line; see Fig. 2-12.

3. Connect all the points on line U. Connect all the points on line L. Each pair of points in step 2b or 2c is plotted at x_r, the value of the random variable at sample-point

Example No. 11

Use the relay data in Example No. 10 in par. 2-4.1. See if they might reasonably have come from an exponential distribution, i. e., $F_{hyp}(x) = 1 - \exp(-x/\theta)$. For simplicity of notation, define $F_i = F_{hyp}(x_i)$.

Procedure	Example
1. Calculate the sample mean.	1. Sample mean = 16,994 = θ.
2. Find the critical value of d from Table 2-12; 90% s-confidence seems reasonable.	2. For $N = 50$, and s-confidence = 90% we have $d = 0.17$.
3. Since we estimated the exponential parameter from the data, multiply d by 0.8.	3. New $d = 0.17 \times 0.8 = 0.14$.
4. Begin with lowest value, $i = 1$. (F_i denotes the actual Cdf at sample-point i.)	4. $F_1 = 1 - \exp(-$ first failure/sample mean) $F_1 = 1 - \exp(-1283/16994) = 0.073$. $0/50 = 0.00$; $1/50 = 0.02$. $d_1 = 0.073 < 0.14$; OK.
5a. Calculate i_A and i_B from Eqs. 2-46 and 2-47. Use the i to denote the actual intersections in Fig. 2-12(B). They will generally be fractional values.	5a. $i_0 = 1$ $(i/50) - 0.14 = F_1 = 0.073$ $i = 10^+$, $i_A = 11$ $F_{i_B} \geq (1/50) + 0.14 = 0.16$ $F(x_i) = 0.16$ $0.16 = 1 - \exp(-x_i/16994)$ $x_i = 2962$, $i_B = 5$.
b. Find the next $i_0 = \min\{i_A, i_B\}$; evaluate F_i. Compare with i_0/N and with $(i_0 - 1)/N$.	b. $i_0 = \min\{5, 11\} = 5$, $x_5 = 3437$. $F_5 = 0.183$ (by same procedure as in step 4) $i_0/N = 5/50 = 0.10$; $(i_0 - 1)/N = 0.08$. The maximum difference is $d_5 = 0.103$ which is less than 0.14; OK.

Example No. **11** (Cont'd)

6a. Repeat step 5a with new i_0. -

6a. $(i/50) - 0.14 = F = 0.183$

$i = 16^+, \; i_A = 17$

$F_{i_B} \geq (5/50) + 0.14 = 0.24$

$x_i = 4663, \; i_B = 11$ (by same procedure as in step 5)

b. Repeat step 5b with new i_0.

b. $i_0 = \min\{11, 17\} = 11, x_{11} = 4865$

$F_{11} = 0.249$

$i_0/N = 11/50 = 0.22; \; (i_0 - 1)/N = 0.20.$
$d_{11} = 0.049 < 0.14;$ OK.

7a. Repeat step 5 with new i_0.

7a. $(i/50) - 0.14 = F_{11} = 0.249, \; i = 19^+, \; i_A = 20$

$F_{i_B} \geq (11/50) + 0.14 = 0.36$

$x_i = 7584, \; i_B = 19.$

b. Repeat step 5b with new i_0.

b. $i_0 = \min\{19, 20\} = 19, \; x_{19} = 8025$

$F_{19} = 0.376$

$i_0/N = 19/50 = 0.38; \; (i_0 - 1)/N = 0.36.$
$d_{19} = 0.016 < 0.14;$ OK.

8a. Repeat step 5a.

8a. $(i/50) - 0.14 = F_{19} = 0.376, \; i_A = 26.$

$F_{i_B} \geq (19/50) + 0.14 = 0.52$

$x_i = 12{,}473, \; i_B = 29.$

b. Repeat step 5b.

b. $i_0 = 26, \; x_{26} = 9595$

$F_{26} = 0.431$

$i_0/N = 26/50 = 0.52; \; (i_0 - 1)/N = 0.50.$
$d_{26} = 0.089 < 0.14;$ OK.

9a. Repeat step 5a.

9a. $(i/50) - 0.14 = F_{26} = 0.431, \; i_A = 29$

$\geq (26/50) + 0.14 = 0.66$

$x_i = 18{,}333, \; i_B = 38.$

b. Repeat step 5b.

b. $i_0 = 29, \; x_{29} = 12{,}937$

$F_{29} = 0.533$

$i_0/N = 29/50 = 0.58; \; (i_0 - 1)/N = 0.56.$
$d_{29} = 0.047 < 0.14;$ OK.

Example No. **11** (Cont'd)

10a. Repeat step 5a.

10a. $(i/50) - 0.14 = F_{29} = 0.533$, $i_A = 34$

$F_{i_B} \geq 29/50 + 0.14 = 0.72$

$x_i = 21{,}632$, $i_B = 39$.

b. Repeat step 5b.

b. $i_0 = 34$, $x_{34} = 14{,}988$

$F34 = 0.586$

$i_0/N = 34/50 = 0.68$; $(i_0 - 1)/N = 0.66$.
$d34 = 0.094 < 0.14$; OK.

11a. Repeat step 5 with new i_0.

11a. $(i/50) - 0.14 = F34 = 0.586$, $i_A = 37$

$F_{i_B} \geq 34/50 + 0.14 = 0.82$

$x_i = 29{,}141$; $i_B = 42$

b. Repeat step 5b.

b. $i_0 = 37$, $x_{37} = 17{,}807$

$F37 = 0.649$

$37/50 = 0.74$; $36/50 = 0.72$.
$d37 = 0.091 < 0.14$; OK.

12. Repeat step 5.

12. $(i/50) - 0.14 = F37 = 0.649$, $i_A = 40$

$F_{i_B} \geq 37/50 + 0.14 = 0.88$

$x_i = 36{,}031$; $i_B = 44$

$i_0 = 40$, $x_{40} = 23{,}449$

$F_{40} = 0.748$

$40/50 = 0.80$; $39/50 = 0.78$.
$d_{40} = 0.052 < 0.14$; OK.

13. Repeat step 5.

13. $(i/50) - 0.14 = F_{40} = 0.748$, $i_A = 45$

$F_{i_B} \geq 40/50 + 0.14 = 0.94$

$x_i = 47{,}811$; $i_B = 47$

$i_0 = 45$, $x_{45} = 41{,}554$

$F_{45} = 0.913$

$47/50 = 0.94$; $46/50 = 0.92$.
$d_{45} = 0.027 < 0.14$; OK.

Example No. 11 (Cont'd)

14, Repeat step 5.

14. $(i/50) - 0.14 = F_{45} = 0.913, \ i_A > 50$

$F_{i_B} \geq 45/50 + 0.14 = 1.04, \ i_B > 50.$

Thus x_{50} must be OK and the distribution i not rejected.

Only 11 trials (i = 1, 5, 11, 19, 26, 29, 34, 37, 40, 45, 50) were necessary **for** the 50 points. Interestingly enough, we now have two distributions to explain the relay data in par. **2-4.1.** Unless there are compelling reasons to the contrary, one is rightly tempted to pick the simpler distribution.

number *r*. The region between lines U and L is the s-confidence envelope; i.e., the true *Cdf* lies entirely within the U, L-envelope with the s-confidence level chosen in step 1. The larger the s-confidence level, the wider the envelope and the less informative is the conclusion to be drawn.

If the plot is made on special graph paper such that the desired distribution is a straight line, then a distribution of that form is completely acceptable (with s-confidence *C*) if a straight line fits within the envelope.

Of course, as usual, no guarantees are made for extrapolations outside the range of the data. If extrapolations are made, be sure to show the uncertainty range (this presumes that the form of the distribution is correct). The actual uncertainty, which includes doubt about the form of the distribution, is usually much greater. But even the calculated range of statistical uncertainty is usually discouraging enough.

2-6 NONPARAMETRIC ESTIMATION

Nonparametric methods can be used to estimate reliability and mean life; i.e., it is not necessary to make any assumptions concerning the time-to-failure distribution (Refs. 3 and 10). Nonparametric reliability applies to the test time interval only and cannot be extrapolated in the time domain. This is the same as estimating a binomial parameter. See par. 2-3.1 for more details. If the failure times are known, a nonparametric s-confidence limit for the *Cdf* can be calculated as shown in par. 2-5. See par. 2-2.2 when censoring occurs. For cases which do not fit the techniques in this handbook, a statistician ought to be consulted; it is very easy to *go* astray.

The simplest way to estimate the unreliability for a time interval, is to calculate the proportion of items that fail over that interval.

2-6.1 MOMENTS

The mean and standard deviation of a population can be estimated by equating them to the sample mean and standard deviation, respectively. This can be done without regard to the actual distribution. This is the way the parameters of the s-normal and exponential distributions are estimated. Weighted sample moments can be used if desired. For example, the logarithm of the mean can be estimated as the mean of the logarithms of the data. Each weighting will give a different answer, but the scatter is probably less than the uncertainty anyway.

Since moments of a population can depend very heavily on the tail regions of the distribution, and since very few (if any) data are collected there, it is usually best to use quantile estimators. Quantile estimators are remarkably insensitive to the actual behavior in tail regions.

If there are contractual obligations concerning nonparametric estimation of moments, a competent statistician ought to be consulted.

2-6.2 QUANTILES

Sample quantiles often are used as estimates of population quantiles. Population quantiles (and combinations thereof) often are used to indicate population characteristics. For example, the median is a good measure of the "central tendency", and the distance between the 75% and 25% points is a good measure of the dispersion.

Point estimates and s-confidence limits can be obtained using the method in par. 2-5. Par. 2-2.1 also contains material on point estimates and s-confidence limits.

If the sample has been severely censored, it may not be feasible to use these methods.

In this situation, a competent statistician ought to be consulted. One may even wish to consult the statistician before planning the experiments or trying to get historical data.

2-7 ANALYSIS OF VARIANCE

During system design and development, it may be desirable to establish the relationship between reliability and specific environmental parameters (Refs. **2** and *6).* Also, the system designers may wish to determine if changes in environmental factors or combinations of them have an important impact on reliability. It is possible, by careful experimental design, to obtain a considerable amount of information, even with smaller sample sizes. Two techniques will be discussed—the analysis of variance and regression analysis.

Analysis of variance permits the effects of individual or combinations of several environmental factors on reliability to be determined. By use of regression analysis, an equation can be derived which relates reliability to environmental parameters.

It is very difficult to design a test that covers the entire range of environments that an equipment experiences in practice (Refs. *6* and 16). However, if the most important parameters can be isolated, a test can be designed around them alone with all the other parameters being ignored. During the test, the test parameters are allowed to assume a range of values which simulate the operational environment of the equipment. The behavior of the test units under various combinations of the test parameters is observed. It is then possible to use analysis of variance to determine the effect of each parameter acting **singly** or in combination.

The ability to analyze the effects of combinations of parameters is a very useful part of analysis of variance. A component may be reliable at a certain level of temperature. It may also be reliable at a certain vibration level. However, a combination of these same environmental levels may cause serious degradation in reliability. Analysis of variance permits the effects of these interactions to be evaluated.

Three sources of variations in reliability are considered: (1) variations caused by each environment acting singly, (2) variations caused by combinations of environments, and **(3)** a remainder (the residual error) which is caused by slight variations in the production processes and test equipment fluctuations. The residual error is used as a standard against which the other sources of variability are compared to determine their statistical significance.

2-7.1 STATISTICAL EXPLANATIONS

Just as the name implies, Analysis of Variance (ANOVA) analyzes the variances of a set of data to see if some effects are real, or just due to random sampling effects. Three categories for measurements are involved:

1. Factors (e.g., heat treatment, supply voltage, humidity)

2. Levels within a factor (e.g., high voltage, usual voltage, low voltage)

3. Replication within levels (e.g., 10 measurements for each voltage).

Each measurement is of a performance characteristic such as strength or time-to-failure.

Just to get a broad picture of what is involved, consider the following experiment on **some** radio receivers. The time-to-first-failure is to be measured for each receiver.

1. ***Factors.*** There are 2 factors:

a. Ambient temperature

b. Supply voltage

2. Levels. There are 2 levels of ambient temperature:

a. High

b. Usual

and 3 levels of supply voltage

a. High

b. Usual

c. Low

3. Replication. There are **4** receivers operated at each possible condition.

Table 2-14 shows the measurements which will be made. This is called a full-factorial experiment since all possible combinations of levels and factors occur. Full-factorial experiments are often too expensive and time consuming to run. Suppose that:

1. The average life of a receiver under the experimental conditions is 3 months.

2. There are 6 test stands, each costing $10,000.

3. Time on a test stand costs $1000 per month.

It is readily seen that the experiment will consume a calendar year and will cost $132,000 just to set it up and run it—assuming nothing goes wrong.

Fractional factorial experiments are discussed in texts on experimental design. This is a sophisticated subject and requires a knowledge of statistics and engineering. Only full factorial designs are considered in the remainder of this paragraph.

The assumption is made in analyzing the data that all measurements are actually from the same population, i.e., factors and levels have absolutely no effect; this is the null hypothesis. It is foreseen, however, that the mean lives under certain conditions may not be the same; this is then taken into account in calculating the various sample variances. Then we see, according **to** the actual data, how likely we were to get the results we got. If the results would be very unlikely—say less than 1 chance out of 1000—we usually then reject the original assumption. The details of the analysis are more complicated, of course.

As the name implies, we estimate the variance of the data in several ways, and then compare the variance estimates. In order to make the comparison feasible, the variances must be estimated in a particular way. All of the estimation and analysis can be performed without the assumption of s-normality, up to the point of making quantitative s-significance statements.

Suppose a population has a variance σ^2

TABLE 2-14

EXPERIMENT ON RADIO RECEIVERS

Each x represents a measurement of life.

		Supply Voltage j		
		Low	Usual	High
Ambient Temperature i	Usual	x	x	x
		x	x	x
		x	x	x
		x	x	x
	High	x	x	x
		x	x	x
		x	x	x
		x	x	x

There are 24 (2 X 3 X '4 = 24) measurements to be made.

and mean μ (the distribution need not be s-normal). Consider samples of size N, and the means x_i of those samples. The x_i will have variance σ^2/N and mean μ. For example, in Table 2-14, if we take the mean of each sample of **4** (N = **4**) in each box, those means will have a variance of $\sigma^2/4$. This fact, in the analysis of variance, usually is used in reverse: if the variance of those means is σ_4^2, then $\sigma^2 = 4\sigma_4^2$. That is, we multiply the "variance of the means" by the "sample size" to get the original variance.

When estimating a population variance from a sample variance, the phrase degrees-of-freedom often is used. The degrees-of-freedom for a sample is the number of s-independent measurements in the sample. In the usual case of a simple sample of N items, the sample mean is subtracted from each measurement: the sum of these deviations is zero. Thus only $N - 1$ are s-independent; once you know those, the last one is uniquely determined. In the analysis of variance manipulations, calculating the degrees-of-freedom is more complicated because there are many subsample means used in the calculations.

The big trick in making estimates of the population variance is to find a set of measurements that are s-independent. There are many ways to estimate the variance, only a few of them are useful. After a brief discussion of notation, the useful ways of estimating the population variance will be shown.

Use the following notation:

x = coded experimental value. It has subscripts. All values of x are measured from the overall mean; i.e., the overall mean of the data has been subtracted from the original experimental value. This simplifies the equations.

i, j, k, r = designates the factor or replication.

i, j, k = subscripts. i refers to the level of the first factor, j to the level of the second, k to the level of the third. The unused ones are omitted when appropriate.

r = subscript for replication. (It is omitted when not used.) It follows the i, j, k.

$\bar{x}$ = a mean value of x. A dot is used to replace the index which has been averaged over.

I, J, K = number of levels assumed by factors **i, j, k,** respectively.

R = number of replications. If r is omitted, it is equivalent to R = 1; e.g., 1 replication means 1 measurement, **3** replications (R = 3) means **3** measurements.

Σ_ϕ = implies a sum over the index ϕ from 1 to Φ.

SS = sum of squares; the subscripts **i, j, k, r** show what variables the sum is due to.

ν = degrees-of-freedom; the subscripts **i, j, k, r** show what variables are being referred to.

s^2 = variance estimate; the subscripts i, j, **k, r** show what variables the estimate is due to

In Table **2-14,** ambient temperature is the first factor **(i)** and I = 2; supply voltage is the second factor (j) and J = **3;** there is no third factor; there are **4** replications and R =

4. The general variable is x_{ijr}; $\bar{x}_{ij\cdot}$ means an average (in cell i, j) over all replications; $x\ldots$ means the average over everything and is zero by definition (see x).

2-7.2 CASE I: 1 FACTOR, WITH REPLICATION, TABLE 2-15

x_{ir} is the variable. It is subdivided as shown in Eq. 2-49.

$$x_{ir} = (\bar{x}_{i\cdot}) + (x_{ir} - \bar{x}_{i\cdot}) \qquad (2\text{-}49)$$

Each of the terms in ( ) in Eq. 2-49 is used to estimate the variance.

$$\Sigma_i \Sigma_r x_{ir}^2 = \Sigma_i \Sigma_r (\bar{x}_{i\cdot})^2 + \Sigma_i \Sigma_r (x_{ir} - \bar{x}_{i\cdot})^2 \qquad (2\text{-}50)$$

TABLE 2-15

CASE I: 1 FACTOR, WITH REPLICATION

(A) GENERAL CASE

Factor i

		1	2	...	I
Replication	1	x	x		x
	2	x	x		x
	3	x	x	...	x
	.	.	.		.
	.	.	.		.
	R	x	x		x

(B) EXAMPLE, 1 FACTOR, 3 LEVELS, 4 REPLICATIONS

Factor 1

Level:	1	2	3
	2.98	−0.47	0.04
	−0.39	1.24	−0.59
	−0.01	−0.40	−0.51
	−0.23	−1.69	0.03
Mean	0.588	−0.330	−0.258

Entries adjusted so that grand average is zero.

Eq. 2-50 uses the fact that the cross products vanish due to wise choice of expressions in the ( ); this will be true in all cases.

The allocation of sums of squares and degrees of freedom is most easily visualized in a table. This case is shown in Table 2-16(A). The residuals estimate of the variance is used as a reference; it would give the common within-factor variance even if the factors were causing a shift in the means. The ratio s_i^2/s_r^2 is tested by means of the F statistic—the distribution of the ratio of 2 s-independent s^2 from the same population. If the ratio is very high (rarely exceeded) it is doubtful that the s_i^2 measures only the common variance; there is very probably a real difference in the means.

Example No. 12 illustrates the case of one factor with replication.

The data in Table 2-15(B) were actually all taken from the table of standard s-normal random deviates (zero mean, unit variance) in Ref. 25 (p. 396). If we had found

TABLE 2-16

ALLOCATION OF SUMS OF SQUARES AND DEGREES OF FREEDOM

(A) Calculations and Allocations for Case I
(The overall mean is presumed to be zero).

Allocation	SS	ν	s^2	F
i	$\Sigma_i \Sigma_r (\bar{x}_{i\cdot})^2$	$I-1$	SS_i/ν_i	s_i^2/s_r^2
	$R\Sigma_i (\bar{x}_{i\cdot})^2$			
r	$\Sigma_i \Sigma_r (x_{ir} - \bar{x}_{i\cdot})^2$	$I(R-1)$	SS_r/ν_r	–
Total	$\Sigma_i \Sigma_r x^2{}_{ir}$	$IR-1$	–	–

(B) Example for Case I: $I = 3$, $R = 4$

Allocation	SS	ν	s^2	F
i	2.0848	2	1.0424	0.7572
r	12.3894	9	1.3766	–
Total	14.4708	11	–	–

(The total does not agree with $SS_i + SS_r$ because of roundoff errors.)

Example No. 12

Table 2-15(B) shows some simulated experimental data. There is 1 factor i with 3 levels. The question is, do those 3 levels actually have a real effect? For example, level 1 appears to have a much higher mean than do the other 2 levels.

Procedure	Example
1. Subtract the overall mean from each datum.	1. This has already been done in Table 2-15(B).
2. Find the mean of each column.	2. See Table 2-15(B).
3. The column means have a dispersion. Estimate the variance of the column means.	3. The mean of the "column means" is zero; so the sum of squares is $(0.588)^2 + (-0.330)^2 + (-0.258)^2 = 0.5212$. There are 3 columns with known mean; so there are $\nu_i = 2$ $(2 = 3 - 1)$ degrees of freedom (for the numerator in step 6): $0.5212/2 = 0.2606$. This is the estimated variance of the column means.
4. Estimate the population variance from the column-mean variance.	4. To convert it to the estimated population variance, multiply by the number of elements in each column (4). Therefore $s_i^2 = 4 \times 0.2606 = 1.0424$.
5. Estimate the variance from the replications. First, get the sample (column) sums of squares. Then calculate the degrees of freedom. Calculate the s_r^2. This is an estimate of the population variance because it was derived from individual data, not the means.	5. Column 1 sum of squares is $(2.98 - 0.588)^2 + (-0.39 - 0.588)^2 + (-0.01 - 0.588)^2 + (-0.23 - 0.588)^2 = 7.7049$. Column 2 sum of squares is 4.3390. Column 3 sum of squares is 0.3455. The total replication sum of squares is $7.7049 + 4.3390 + 0.3455 = 12.3894$. Each column has 3 $(3 = 4 - 1)$ degrees of freedom and there are 3 columns. There are $3 \times 3 = 9$ degrees of freedom for the sum of squares; therefore $\nu_r = 9$ for step 6. $s_r^2 = 12.3894/9 = 1.3766$.

Example No. 12 (Contfd)

6. Compare s_r^2 and s_i^2. The technique for doing this is the F distribution.	6. $s_i^2/s_r^2 = 1.0424/1.3766 = 0.7572$ $\nu_i = 2$, $\nu_r = 0$ from steps 3 and 5. Look in the F tables (Tables 2-17) for $\nu = 2$ in the numerator and $\nu = 9$ in the denominator. The critical value for ,s-significance at the 1% level is given in Table 2-17 and is 8.02. The ratio 0.7572 is much less than the critical value; so we presume that the null hypothesis is true, i. e., all data are random samples from a single population, there is no real difference due to the levels of the factor. The actual value of F corresponding to 0.7572 is 50^{+}% (see Tables in Ref. 13).

TABLE 2-17

F-DISTRIBUTION, $F_{.99}$ (ν_n, ν_d) (Ref. 18)

Critical values for the 1% (1% = 100% − 99%) s-significance level.

Table gives the value of F which is exceeded only 1% of the time.

$$F \equiv s_n^2 / s_d^2$$

ν_n = degrees of freedom for numerator; ν_d = degrees of freedom for denominator

ν_d \ ν_n	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.91	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.18	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	3.67	5.32	5.07	4.59	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.65	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.05	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.81	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

ν_d = degrees of freedom for denominator

(The F-distribution is sometimes called the ν^2 distribution; it is the ratio of 2 s-independent, s-unbiased estimates from a single s-normal distribution.)

an effect, it would have been an erroneous finding.

This is about the simplest possible case. There are several algebra identities for reducing the amount of arithmetic (see Ref. 32, for example); their use may incur rounding errors which can be severe. An example of rounding errors can be shown using the identity

$$y = (a + b)(a - b) = a^2 - b^2 \quad (2\text{-}51)$$

Suppose a = 100,003, b = 100,002 and the computer has only 6 significant figures. If the factored formula is used, we have

(100,003 + 100,002) (100,003 – 100,002) = 200,005 X 1 = 200,005.

If the other formula is used, we have

$100{,}003^2 - 100{,}002^2$ = 100,006 X 10^5 – 100,004 X 10^5 = 200,000.

The rounding error by the second method caused a loss of almost 1 significant digit.

Another advantage of not using "short cut" methods is that one has a chance to see the residuals and to note any that may be anomalous.

Most analyses actually will be done via a computer program. Check your computer service center to find out which ones are available to you.

2-7.3 CASE II: 2 FACTORS, WITHOUT REPLICATION, TABLE 2-18

It is usually a poor idea to have no replication. Without replication, one must resort to precarious assumptions to estimate the reference for the population variance.

x_{ij} is the variable. It is subdivided as shown in Eq. 2-52 ($\bar{x}$ . . . = 0).

$$x_{ij} \equiv (\bar{x}_{i\cdot}) + (\bar{x}_{\cdot j}) + (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j}) \quad (2\text{-}52)$$

It can be shown that, as in Table 2-18(A), the sums of cross products vanish. The difficulty with this case (no replication) is that there is presumed to be no interaction; therefore the s_{ij}^2 is taken as the reference for the F-ratio test. · Table 2-18(B) shows graphically what tests are run.

2-7.4 CASE III: 2 FACTORS, WITH REPLICATION, TABLE 2-19

x_{ijr} is the variable. It is subdivided as shown in Eq. 2-53.

$$x_{ijr} = (\bar{x}_{i\cdot\cdot}) + (\bar{x}_{\cdot j\cdot}) + (\bar{x}_{ij\cdot} - \bar{x}_{i\cdot\cdot} - \bar{x}_{\cdot j\cdot}) + (x_{ijr} - \bar{x}_{ij\cdot}) \quad (2\text{-}53)$$

It can be shown that, as in Table 2-19(A), the sums of cross products vanish. The first 2 terms will give the main effects of the 2 factors i and j; the 3rd term gives the interaction effect; and the last term gives the estimate of the population variance from the replication (this is used as the reference for testing the other effects). Table 2-19(B) shows the experimental layout; there are *IJR* experiments to be run. Table 2-19(A) shows that there are 3 F-tests. If the s-significance level is too loose, there is a good chance that one of the 3 effects will be declared s-significant when the F-value is high just due to chance.

2-7.5 CASE IV: 3 FACTORS, WITHOUT REPLICATION, TABLE 2-20

As in Case II, it is usually a poor idea to have no replication; but the number of tests with replication can be prohibitive. x_{ijk} is the variable. It is subdivided as shown in Eq. 2-54. ($\bar{x}$ . . . = 0).

TABLE 2-18

CASE II: ANALYSIS VARIANCE, 2 FACTORS, WITHOUT REPLICATION

(A) Calculation and Allocations
(The overall mean $\bar{x}$. . is presumed to be zero.)

Allocation	SS	ν	s^2	F
i	$\Sigma_i \Sigma_j (\bar{x}_{i.})^2$	$I - 1$	SS_i/ν_i	s_i^2/s_{ij}^2
j	$\Sigma_i \Sigma_j (\bar{x}_{.j})^2$	$J - 1$	SS_j/ν_j	s_j^2/s_{ij}^2
ij	$\Sigma_i \Sigma_j (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j})^2$	$(I-1)(J-1)$	SS_{ij}/ν_{ij}	–
Total	$\Sigma_i \Sigma_j (x_{ij})^2$	$IJ - 1$	–	–

The ij implies an interaction term.

(B) Experimental Layout
Each x represents an experiment.

Factor i

	j i	1	2	...	I
	1	x	x	...	x
Factor j	2	x	x	...	x
	J	x	x	...	x

$$
\begin{aligned}
x_{ijk} \equiv\ & (\bar{x}_{i..}) + (\bar{x}_{.j.}) + (\bar{x}_{..k}) \\
& + (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.}) \\
& + (\bar{x}_{.jk} - \bar{x}_{.j.} - \bar{x}_{..k}) \\
& + (\bar{x}_{i.k} - \bar{x}_{..k} - \bar{x}_{i..}) \\
& + [x_{ijk} - (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.}) \\
& - (\bar{x}_{.jk} - \bar{x}_{.j.} - \bar{x}_{..k}) \\
& - (\bar{x}_{i.k} - \bar{x}_{..k} - \bar{x}_{i..}) \\
& - (\bar{x}_{i..}) - (\bar{x}_{.j.}) - (\bar{x}_{..k})] \qquad (2\text{-}54)
\end{aligned}
$$

The first 3 terms give the main effects due to the 3 factors; the second 3 terms give the 2-way interaction effects; and the last term in [ ] gives the 3-way interaction term. The [. . .] is written out in detail to show how it is constructed; it can be simplified in appearance somewhat as in Table 2-20(A). Whenever a term is written down, all further averages must be subtracted from it; in the end, the identity must also be preserved.

a. The first 3 terms have no further averages because $\bar{x}$. . . = 0 (by hypothesis; if not, $\bar{x}$ . . . is subtracted from every reading).

b. Consider the first 2-way interaction. The $\bar{x}_{ij}$. has 2 further averages $\bar{x}_i$ . . and $\bar{x}_{.j.}$; each must be subtracted from $\bar{x}_{ij}$. But the $\bar{x}_i$. . and $\bar{x}_{.j}$. have no further (non-zero) averages because $\bar{x}$. . . = 0. The same considerations hold for the other 2-way interactions.

c. The 3-way interaction term begins with x_{ijk}. It has 3 further averages (over each of the indexes); each of those averages has 2 further averages—and just repeats the 2-way interaction terms; so each of those is sub-

TABLE 2-19

CASE III: ANALYSIS OF VARIANCE, 2 FACTORS WITH REPLICATION

(A) Calculations and Allocations

(The overall mean $\bar{x}\ldots$ is presumed to be zero.)

Allocation	SS	ν	s^2	F
i	$\Sigma_i\Sigma_j\Sigma_r(\bar{x}_{i\cdot\cdot})^2$	$I-1$	SS_i/ν_i	s_i^2/s_r^2
j	$\Sigma_i\Sigma_j\Sigma_r(\bar{x}_{\cdot j\cdot})^2$	$J-1$	SS_j/ν_j	s_j^2/s_r^2
ij	$\Sigma_i\Sigma_j\Sigma_r(\bar{x}_{ij\cdot}-\bar{x}_{i\cdot\cdot}-\bar{x}_{\cdot j\cdot})^2$	$(I-1)(J-1)$	SS_{ij}/ν_{ij}	s_{ij}^2/s_r^2
r	$\Sigma_i\Sigma_j\Sigma_r(x_{ijr}-\bar{x}_{ij\cdot})^2$	$IJ(R-1)$	SS_r/ν_r	—
Total	$\Sigma_i\Sigma_j\Sigma_r(x_{ijr})^2$	$IJR-1$	—	—

(B) Experimental Layout

Each x represents an experiment.

Factor i

	$j \backslash i$		1	2	—	I
Factor j	1	1	x	x		x
		2	x	x		x
					. . .	
			.	.		.
		R	x	x		x
	2	1	x	x		x
		2	x	x		x
					. . .	
			.	.		.
		R	x	x		x
	.	.	.	.	.	.
	.	.	.	.	.	.
		.	.	.	.	.
	J	1	x	x		x
		2	x	x		x
					. . .	
			.	.		.
		R	x	x		x

tracted. Then there are the 3 double averages $(\bar{x}_{i\cdot\cdot})$ etc. which must be subtracted; if $\bar{x}\ldots$ were not zero, it would have to be subtracted every time. Then the triple average $\bar{x}\ldots$ would have to be subtracted if it were not zero. (This shows what simplification is achieved by making "$\bar{x}\ldots = 0$"; it also virtually eliminates roundoff errors in computer calculations.

Table 2-20 shows the equations and experimental layout. Example No. 13 illustrates the case of three factors without replication.

[Text continues on page 2-94.]

TABLE 2-20

CASE IV: ANALYSIS OF VARIANCE, 3 FACTORS WITHOUT REPLICATION

(A) Calculations and allocations

(The overall mean $\bar{x}_{...}$ is presumed to be zero.)

allocation	SS	ν	s^2	F
i	$\Sigma_i\Sigma_j\Sigma_k(\bar{x}_{i..})^2$	$I-1$	SS_i/ν_i	s_i^2/s_{ijk}^2
j	$\Sigma_i\Sigma_j\Sigma_k(x_{.j.})^2$	$J-1$	SS_j/ν_j	s_j^2/s^2_{ijk}
k	$\Sigma_i\Sigma_j\Sigma_k(x_{..k})^2$	$K-1$	SS_k/ν_k	s_k^2/s_{ijk}^2
ij	$\Sigma_i\Sigma_j\Sigma_k(\bar{x}_{ij.}-\bar{x}_{i..}-\bar{x}_{.j.})^2$	$(I-1)(J-1)$	SS_{ij}/ν_{ij}	s_{ij}^2/s_{ijk}^2
jk	$\Sigma_i\Sigma_j\Sigma_k(\bar{x}_{.jk}-\bar{x}_{.j.}-\bar{x}_{..k})^2$	$(J-1)(K-1)$	SS_{jk}/ν_{jk}	s_{jk}^2/s_{ijk}^2
ki	$\Sigma_i\Sigma_j\Sigma_k(\bar{x}_{i.k}-\bar{x}_{..k}-\bar{x}_{i..})^2$	$(K-1)(I-1)$	SS_{ki}/ν_{ki}	s_{ki}^2/s_{ijk}^2
ijk	$\Sigma_i\Sigma_j\Sigma_k(\bar{x}_{ijk}-\bar{x}_{ij.}-\bar{x}_{i.k}-\bar{x}_{.jk}+\bar{x}_{i..}+\bar{x}_{.j.}+\bar{x}_{..k})^2$	$(I-1)(J-1)(K-1)$	SS_{ijk}/ν_{ijk}	–
[illegible]	[illegible]	$IJK-1$		

(B) Experimental Layout

(Each x represents an experiment.)

				Factor i			
	i			1	2	. . .	
Factor j	1	Factor k	1	x	x		x
			2	x	x	. . .	x
					. . .		
			K	x	x		x
	2	Factor k	1	x	x		x
			2	x	x	. . .	x
			K	x	x		x
	J	Cactor k	1	x	x		x
			2	x	x	. . .	x
			K	x	x		x

Example No. 13

There are serious reliability difficulties with the gun/turret drive-system on a heavy tank. A full factorial experiment with no replication is to be run. Factor **t** is temperature and there are 2 levels; factor **v** is vibration and there are 3 levels; factor **h** is humidity and there are 2 levels. There are 12, i. e., $(2 \times 3 \times 2 = 12)$, experiments. Table 2-21(B) shows the data for this case. Each datum is the number of 10-hr missions-to-failure.

Procedure	Example
1. Record the data. Convert to "overall mean is zero." Begin to fill out a table.	1. Table 2-21(B) is the original data. Subtract 5.25 from each to get Table 2-21(C). Begin Table 2-21(A) (patterned after Table 2-20(A)). Cols. 1 and 3 can be completed.
2. Calculate overall averages for the temperature factor.	2. $\bar{x}_{1..} = (4.75 + 3.75 + 2.75 + 3.75 - 0.25 - 0.25)/6 = 2.4167$ $\bar{x}_{2..} = -2.4167.$
3. Calculate overall averages for the other factors	3. $\bar{x}_{.1.} = (4.75 - 1.25 + 3.75 - 0.25)/4 = 1.7500$, $\bar{x}_{.2.} = 0.5000$, $\bar{x}_{.3.} = -2.2500$; $\bar{x}_{..1} = (4.75 - 1.25 + 2.75 - 2.25)/4 = -0.0833$, $\bar{x}_{..2} = 0.0833.$
4. As a check, the sum of the factor averages, for each factor, must be zero.	4. OK, by inspection.
5. Calculate the first 3 SS's: t, v, h. The sums shown in Table 2-20(A) can be simplified because, in each case, 2 sums are trivial: they can be replaced with a multiplication. That multiplication is also exactly what is needed to convert the variance of the means to a variance of the population.	5. $SS_t = 3 \times 2 \times [(2.4167)^2 + (-2.4167)^2] = 6 \times 11.6809 = 70.0853$ SS, $= 2 \times 2 \times 8.3750 = 33.5000$ $SS_h = 3 \times 2 \times 0{,}0139 = 0.0833$ Record in Table 2-21(A), Col. 2.
6. Calculate $s^2 = SS/\nu$ for each of the 3 factors.	6. $s_t^2 = 70.0853/1 = 70.0853$ $s_v^2 = 16.7500$ $s_h^2 = 0.0833$; Record in Table 2-21(A) Col. 4.

Example No. 13 (Cont'd)

7. For the **tv** interaction, calculate the $\bar{x}_{tv\cdot}$.

7. $\bar{x}_{11\cdot} = (4.75 + 3.75)/2 = 4.25$

$\bar{x}_{12\cdot} = (2.75 + 3.75)/2 = 3.25$

$\bar{x}_{13\cdot} = (-0.25 - 0.25)/2 = -0.25$

$\bar{x}_{21\cdot} = -0.75$, $x_{22\cdot} = -2.25$, $x_{23\cdot} = -4.25$.

8. Calculate the t,v terms and the sum of squares. Use the results from steps 2, 3. Calculate the s^2.

8. $(4.25 - 2.4167 - 1.7500)^2 = (0.0833)^2$

$(3.25 - 2.4167 - 0.5000)^2 = (0.3333)^2$

$(-0.25 - 2.4167 + 2.2500)^2 = (-0.4167)^2$

$(-0.75 + 2.4167 - 1.7500)^2 = (-0.0833)^2$

$(-2.25 + 2.4167 - 0.5000)^2 = (-0.3333)^2$

$(-4.25 + 2.4167 + 2.2500)^2 = \underline{(0.4167)^2}$

Total 0.5833

Multiply by 2 ($H = 2$) and record in Table 2-21(A) Col. 2. Divide result by 2 ($\nu_{tv} = 2$) and record in Table 2-21(A), Col. 4.

9. For the **vh** interaction, calculate the $\bar{x}$

9. $\bar{x}_{\cdot 11} = (4.75 - 1.25)/2 = 1.75$

$\bar{x}_{\cdot 12} = 1.75$, $\bar{x}_{\cdot 21} = 0.25$, $\bar{x}_{\cdot 22} = 0.75$

$\bar{x}_{\cdot 31} = -2.25$, $\bar{x}_{\cdot 32} = -2.25$.

10. Calculate the v,h terms, the sum of squares, and the s^2.

10. $(1.75 - 1.7500 + 0.0833)^2 = (0.0833)^2$

$(1.75 - 1.7500 - 0.0833)^2 = (-0.0833)^2$

$(0.25 - 0.5000 + 0.0833)^2 = (-0.1667)^2$

$(0.75 - 0.5000 - 0.0833)^2 = (0.1667)^2$

$(-2.25 + 2.2500 + 0.0833)^2 = (0.0833)^2$

$(-2.25 + 2.2500 - 0.0833)^2 = \underline{(-0.0833)^2}$

Total 0.0833

Multiply by 2 ($T = 2$) and record in Table 2-21(A), Col. 2. Divide result by 2($\nu_{vh} = 2$) and record in Table 2-21(A), Col. 4.

11. For the ht interactions, calculate the $\bar{x}_{t\cdot h}$.

11. $\bar{x}_{1\cdot 1} = (4.75 + 2.75 - 0.25)/3 = 2.4167$

$\bar{x}_{2\cdot 1} = -2.5833$

$\bar{x}_{1\cdot 2} = 2.4167$, $\bar{x}_{2\cdot 2} = -2.2500$.

Example No. 13 (Cont'd)

12. Calculate the h, t terms, the sum of squares, and the s^2.

12. $(2.4167 + 0.0833 - 2.4167)^2 = (0.0833)^2$

$(-2.5833 + 0.0833 + 2.4167)^2 = (-0.0833)^2$

$(2.4167 - 0.0833 - 2.4167)^2 = (-0.0833)^2$

$(-2.2500 - 0.0833 + 2.4167)^2 = \underline{(0.0834)^2}$

Total 0.0278

Multiply by 3 ($V = 3$) and record in Table 2-21(A), Col, 2. Divide result by 1 ($\nu_{ht} = 1$) and record in Table 2-21(A), Col. 4.

13. Calculate the terms for the tvh interactions (Table 2-20(A), the sum of squares, and the s^2.

13. $(4.75 - 4.25 - 1.75 - 2.4167 + 2.4167 + 1.75 - 0.0833)^2 = (0.4167)^2$: 111

$(-1.25 + 0.75 - 1.75 + 2.5833 - 2.4167 + 1.75 - 0.0833)^2 = (-0.4167)^2$: 211

$(2.75 - 3.25 - 0.25 - 2.4167 + 2.4167 + 0.50 - 0.0833)^2 = (-0.3333)^2$: 121

$(-2.25 + 2.25 - 0.25 + 2.5833 - 2.4167 + 0.50 - 0.0833)^2 = (+0.3333)^2$: 221

$(-0.25 + 0.25 + 2.25 - 2.4167 + 2.4167 - 2.25 - 0.0833)^2 = (-0.0833)^2$: 131

$(-4.25 + 4.25 + 2.25 + 2.5833 - 2.4167 - 2.25 - 0.0833)^2 = (0.0833)^2$: 231

$(3.75 - 4.25 - 1.75 - 2.4167 + 2.4167 + 1.75 + 0.0833)^2 = (-0.4167)^2$: 112

$(-0.25 + 0.75 - 1.75 + 2.25 - 2.4167 + 1.75 + 0.0833)^2 = (+0.4166)^2$: 212

$(3.75 - 3.25 - 0.75 - 2.4167 + 2.4167 + 0.50 + 0.0833)^2 = (0.3333)^2$: 122

$(-2.25 + 2.25 - 0.75 + 2.25 - 2.4167 + 0.50 + 0.0833)^2 = (-0.3334)^2$: 222

$(-0.25 + 0.25 + 2.25 - 2.4167 + 2.4167 - 2.25 + 0.0833)^2 = (0.0833)^2$: 132

$(-4.25 + 4.25 + 2.25 + 2.25 - 2.4167 + 2.25 + 0.0833)^2 = (-0.0834)^2$: 232

Total 1.1667; record in Table 2-21(A), Col. 2. Divide by 2 ($\nu_{tvh} = 2$). Record in Table 2-21(A), Col. 4.

Example No. 13 (Contfd)

14. Look up the critical F-values.	14. Choose 1%-s-significance; use Table 2-17. **Record in Table 2-21(A), Col. 6.**
15. Look up actual s-significance levels. (The s-significance levels are the F-survivor function, i.e., the fraction of the time a value of F is exceeded. Low values of s-significance imply an effect; high ones do not.)	15. Use Ref. 13. Estimate the values. Record in Table 2-21(A), Col. 7.

TABLE 2-21

CASE IV: LIFE OF A DRIVE SYSTEM FOR THE GUN/TURRET ON A HEAVY TANK

(A) Calculations and Allocations

#1 Allocation	#2 SS	#3 ν	#4 s^2	#5 F	#6 F 1% s-signif	#7[2] approx. s-signif %
t[1]	70.0853	1	70.0853	120	98.5	0.8
v	33.500	2	16.7500	28.7	99.0	4
h	0.0833	1	0.0833	0.143	98.5	85
tv	1.1667	2	0.5833	1.00	99.0	50
vh	0.1667	2	0.0833	0.143	99.0	85
ht	0.0833	1	0.0833	0.143	98.5	85
tvh	1.1667	2	0.5833	–	99.0	50
Total[3]	106.2520 106.2500	11	–			

(B) Experimental Results[4]

Factor t

	v	h	t = 1	2
Factor v	1	1	10	4
		2	9	5
	2	1	8	3
		2	9	3
	3	1	5	1
		2	5	1

overall mean = 5.2500
overall s = 3.1079

(C) Modified Experimental Result (overall mean is zero)

v	h	t = 1	2
1	1	4.75	−1.25
	2	3.75	−0.25
2	1	2.75	−2.25
	2	3.75	−2.25
3	1	−0.25	−4.25
	2	−0.25	−4.25

overall mean = 0.0000
overall s = 3.1079

Number of levels

humidity	h,	H = 2
temperature	t,	T = 2
vibration	v,	V = 3

x_{tvh} is the variable.

Notes: (1) The factors are t temperature, h humidity, and v vibration.
(2) Col. 7 is the approximate s-significance for the F-values in Col. 5 (Ref. 13).
(3) The two SS totals differ because of roundoff errors.
(4) Table shows the number of 10-hr missions-to-fail for each combination of factors.

TABLE 2-21 (Cont'd)

(D) Modified (and somewhat controversial) Analysis Calculations and Allocations

#1 Allocation	#2 SS	#3 ν	#4 s^2	#5 F	#6 F % s-signif.	#7 approx. s-signif. %
t	70.0853	1	70.1	190	12.3	<0.05
v	33.500	2	16.8	45.4	9.55	<0.05
h	0.0833	1	0.0833	0.226	12.3	93
all interactions	2.5834	7	0.3691	—	—	—

Now let us analyze the results of Example No. 13:

1. One of the first things to note is that the low ν for the reference s^2 (ν_{tvh} = 2) causes very poor ability to distinguish very high from ordinary ratios of the s^2

2, The temperature effect is undoubtedly important (for the **2** temperatures tested) the vibration is likely to be important; but none of the other effects or interactions is likely to be important.

3. This is a "fixed effects" analysis. The levels are presumed "fixed", not to be a random sample from all possible levels of the factor. See Refs. **31** and 32 or a competent statistician for a fuller discussion of this point.

4. If more tests are to be run, they ought to be on temperature and vibration separately. The separate tests are easier/cheaper to run.

5. Since there was no replication, it was presumed that there was no 3-way interaction. From the looks of the results, it is easy to assume (see conclusion No. *2)* that none of the 2-way interactions is important. Some (but not all, or even most) statisticians would argue that there is now justification for lumping all SS's for all interactions together and to estimate the reference s^2 as (1.1667 + 0.1667 + 0.0833 + 1.1667)/(2 + 2 + 1 + 2) = 0.3691 with 7 degrees of freedom. The new analysis is shown in Table 2-21(D). It declares, more than ever, that temperature and vibration are most important, and that humidity is negligible. Such remanipulations must be treated with caution; consult a competent statistician before basing any important decisions on them.

6. The basic data themselves are very coarse. It is unlikely that they come from a s-normal distribution. **A** lognormal would perhaps be more appropriate. The reference population standard deviation (from Conclusion No. 5) is about 0.61 mission. By going back now and looking at Table 2-21(B) it is not unreasonable that the humidity effect is small. In 3 cells, there was no effect, and in the other cells there was at most 1 mission difference.

7. After running such an expensive test, the small extra cost of several analyses is not unreasonable. The only unreasonable thing would be to place much importance on results at, say, the 5% or 10% s-significance level, because these will occur in 5% to 10% of the calculations.

2-8 REGRESSION AND CORRELATION ANALYSIS

Variations in component part values introduced by manufacturing processes and variations in environmental conditions cause changes in circuit or equipment characteristics which may affect reliability (Refs. 2 and 6). Numerical relationships that relate reliability to design variables can be derived using linear regression analysis.

Regression analysis is a statistical technique that quantitatively defines the "best" fit of a line through a set of data points. Correlation is a related technique but considers the interdependence of the variables rather than the dependence of one on another. The distinction between the two techniques can be rather subtle and Refs. 24 and 32 or other standard works ought to be consulted. Consider leakage current and life of a transistor. It cannot be said that one causes the other (although they may have common causes); so one might be interested in their correlation. But if one is interested in predicting the life by measuring the leakage current, then regression is used. As another example consider the ambient temperature and the life of a transistor. From a physical point of view, it is not useful to speak of life's causing the ambient temperature, only the other way around; so one is interested only in regression,

As used in this chapter, and in many statistical treatments, regression implies linear regression, and correlation implies linear correlation. Further, this chapter considers only 2 variables. For more extensive treatments, see Refs. 12, 17, 24, and 32 or other statistical texts. The problem is to find the "best" straight line for the test data.

In linear regression, Eqs. 2-55 and 2-56 are assumed

$$y = mx' + b' + \sigma\epsilon \tag{2-55}$$

$$\bar{y} = mx' + b' \tag{2-56}$$

where

y = dependent variable (a function of x')

x' = independent variable

m = slope of the line

b' = y-axis intercept of the line

ϵ = standard s-normal variate

σ = a standard deviation of y from $\bar{y}$

$\bar{y}$ = "mean" value of y (a function of x'). It is averaged over ϵ.

i = subscript denoting a particular set of measured values (the ϵ_i is not actually measured, it is there to make an equality) $i = 1, \ldots, N$

Σ_i = implies the sum over all N test data

N = number of test data pairs

"Best" fit is usually defined as the "least-squares" fit, i.e., $\Sigma_i \epsilon_i^2 \rightarrow$ a minimum, by adjusting m and b'. When ϵ is a standard s-normal variate, the maximum likelihood solution implies the least-squares solution.

It is convenient to redefine the x' and b' as follows:

$$a' \equiv (\Sigma_i x'_i)/N$$

$$x \equiv x' - a'$$

$$x_i \equiv x'_i - a'$$

$$b \equiv ma' + b' \tag{2-57}$$

Then the Eqs. 2-55 and 2-56 become

$$y = mx + b + \sigma\epsilon \tag{2-58}$$

$$\bar{y} = mx + b \tag{2-59}$$

The randomness in y can be caused by measuring techniques, experimental uncertainties, and/or component part variations. The least-squares solution is

$$\hat{b} = (\Sigma_i y_i)/N \tag{2-60}$$

$$\hat{m} = (\Sigma_i x_i y_i)/(\Sigma_i x_i^2) \tag{2-61}$$

$$\hat{\sigma}^2 = s^2 = \{\Sigma_i [y_i - (\hat{m}x_i + \hat{b})]^2\}/(N-2)$$

$$= [\Sigma_i (y_i - \hat{b})^2 - \hat{m}^2 \Sigma_i x_i^2]/(N-2) \tag{2-62}$$

The uncertainties in $\hat{b}$, $\hat{m}$, and $\bar{y}^*$ are measured by

$$\text{Var}\{b\} = \sigma^2/N$$

$$\text{Var}\{m\} = \sigma^2/(\Sigma_i x_i^2)$$

$$\text{Cov}\{\hat{m}, \hat{b}\} = 0 \tag{2-63}$$

$$\text{Var}\{\bar{y}^*(x)\} = \text{Var}\{\hat{b}\} + x^2 \text{Var}\{\hat{m}\} \tag{2-64}$$

where $\bar{y}^*$ is a value of $\bar{y}$ to be predicted from x.

Those uncertainties are usually estimated by substituting s^2 for σ^2. Eq. 2-64 is most important because it shows the "lever-arm" effect. For x large enough, the standard deviation of $\bar{y}^*$ is directly proportional to x; this means that the uncertainty in extrapolation can be tremendous. Eq. 2-63 can be used to put s-confidence limits on $\hat{b}$ and $\hat{m}$. Eq. 2-64 gives prediction limits for $\bar{y}^*(x)$. Student t-distribution is usually used.

One ought never to calculate a $\bar{y}^*$ (x) without also estimating its standard deviation by means of Eq. 2-64; it is very often an unpleasant surprise. If, in addition to $\bar{y}^*(x)$, one wishes the uncertainty in y^*, then σ^2 must be added to Var $\{\bar{y}^*\}$.

The linear correlation coefficient ρ between y and x is

$$\rho = \frac{\text{Cov}\{x, y\}}{\sqrt{\text{Var}\{x\}\,\text{Var}\{y\}}} \tag{2-65}$$

It is estimated by

$$\hat{\rho} = \frac{\Sigma_i x_i y_i}{\{[\Sigma_i x_i^2][\Sigma_i (y_i - \hat{b})^2]\}^{1/2}} \tag{2-66a}$$

$$= \hat{m}\left[\frac{\Sigma_i x_i^2}{\Sigma_i (y_i - \hat{b})^2}\right]^{1/2}. \tag{2-66b}$$

$$\hat{\rho}^2 = 1 - \frac{s^2}{\Sigma_i (y_i - \hat{b})^2/(N-2)} \tag{2-66c}$$

The sampling distribution of $\hat{\rho}$ is given in Refs. 18 and **30** and elsewhere. For small sample sizes, the uncertainty in ρ, given $\hat{\rho}$, is distressingly large. For example, if $\rho = 0$ and $N = 10$, then 5% of the time $|\hat{\rho}|$ will be greater than 0.57; or if $\hat{\rho} = 0.5$ and $N = 10$, 95% s-confidence limits on $\hat{\rho}$ are -0.18 and $+0.84$.

Example No. 14 illustrates the procedure.

2-9 ACCEPT/REJECT TESTS—t TEST FOR MEAN OF A s-NORMAL DISTRIBUTION

Tests can verify that an equipment meets a minimum level or they can be used to compare and select the more reliable unit or approach from several alternatives. These tests can be used during system design and development to provide guidance to the engineers who must select among alternate designs (Ref. 18). A typical statement of alternatives is (Ref. 18):

1. There is a difference between the reliability of the units.

2. No difference has been demonstrated.

Another statement of alternatives is:

1. The reliability of equipment A is greater than that of equipment **B.**

2. There is no reason to believe that the reliability of equipment **A** is greater than that of equipment **B.**

Statistical tests are applied to the data and a decision is made between the two alternatives. The hypotheses are selected prior to the test.

Because test information usually is ob-

Example No. 14 (Ref. 2)

A turbojet engine is experiencing blade-fatigue failures due to excessive vibrations at resonance. A test is to be run to determine whether a relationship exists between bench-measured blade resonance points and the actual rpm at which the blade reaches resonance. If this is established, it may be possible, by nondestructive bench measurements, to determine the acceptability of blades for actual engine use, reducing or eliminating this engine failure mode.

Thirty blades are selected at random and tested sequentially. The test consists of building up engine speed and recording the rpm at which resonance occurs and the resonant frequency. The test data are given in Table 2-22. Compute the linear regression equation and the associated statistical quantities.

Procedure	Example
1. Plot the data to see what they look like	1. See Fig. 2-13. They look reasonable enough.
2. Calculate the deviations from the mean for the bench resonant frequencies.	2. Fill in the x column in Table 2-22,
3. Calculate $\Sigma_i x_i^2$.	3. $\Sigma_i x_i^2 = 223{,}266.70$.
4. Calculate $\Sigma_i y_i$.	4. $\Sigma_i y_i = 322{,}200$.
5. Calculate $\Sigma_i x_i y_i$.	5. $\Sigma_i x_i y_i = 2{,}608{,}330$.
6. Calculate $\hat{b}$ and $\hat{m}$ from Eqs. 2-60 and 2-61.	6. $\hat{b} = 10{,}740$ $\hat{m} = 11.683$ $a = 1085.1$.
7. Calculate each ϵ_i, $\epsilon_i = y_i - (\hat{m}x_i + \hat{b})$ then calculate $s^2 = (\Sigma_i \epsilon_i^2)/(N-2)$.	7. $s^2 = 50175.5$ $s = 224.0$.
8. Estimate Var $\{\hat{b}\}$ and Var $\{\hat{m}\}$. Use Eq. 2-63.	8. Var $\{\hat{b}\} \approx 50178.2/30 = 1672.5 = 40.9^2$ Var $\{\hat{m}\} \approx 50178.2/223{,}266{,}7 = 0.2247 = 0.4741^2$

Example No. **15** (Contfd)

9. Write the equation for $\mathrm{Var}\{y^*(x)\}$.

9. $\mathrm{Var}\{y^*(x)\} \approx 1672.5 + 0.2247x^2 + 50175.5$
$= 51{,}848 + 0.2247x^2.$

10. Calculate $\mathrm{Var}\{y^*(x)\}$ for several x.

10. $x = 200$

$\mathrm{Var}\{y^*(200)\} = 51848 + 8988 = 60836 = 247^2$

$\mathrm{Var}\{y^*(100)\} = 54095 = 233^2$

$\mathrm{Var}\{y^*(0)\} = 228^2.$

11. Calculate the linear correlation coefficient. Use Eq. 2-66c.

11. $\Sigma_i(y_i - b)^2 = 31{,}877{,}000$

$\Sigma_i(y_i - b)^2/(N-2) = 31{,}877{,}000/28 = (1066.99)$

$\hat{\rho}^2 = 1 - \left(\frac{224.0}{1066.99}\right)^2 = 0.956$

$\hat{\rho} = 0.978.$

12. Use Table 15 of Ref. **30** to get 95% s-confidence limits for ρ.

12. The chart is not clear for such a high $\hat{\rho}$, but it appears that $\mathrm{Conf}\{0.95 \le \rho \le 0.99) = 95\%$. In any event, the linear correlation is **very** high. Virtually all the original variance in y is explained by the regression.

The statistical data have all been gathered. It now remains to interpret it. With $\nu = 28$, it makes little difference whether Student t-distribution is used, or the s-normal distribution. One could use a goodness-of-fit test on the ϵ_i to see if they could reasonably have come from a s-normal distribution. More disturbing are the very high deviations of blade 6, and the fact that the $|\epsilon|$ appears greater for the half of the sample with higher resonant frequency.

If the engine operating requirements can be satisfied with the present knowledge, no further experimental or statistical tests are needed. But remember, **1/30** of the sample had a deviation exceeding ± **3** standard deviations. Do not blindly forget the sample and then make predictions of the population based on some abstract statistical procedures. Statistics can only answer the questions it is asked. If you don't ask the "right" questions, you will get answers to irrelevant questions.

TABLE 2-22

REGRESSION TEST RESULTS FOR TURBINE BLADES

Blade No.	Resonance x' Frequency, Hz	x	Y rpm	E	Blade No.	Resonance x' Frequency, Hz	x	Y rpm	E
7	960	–125.1	9400	+121.5	27	1062	– 23.1	10550	79.9
18	969	–116.1	9400	+ 16.3	3	1069	– 16.1	10700	148.1
2	986	– 99.1	9550	– 32.3	11	1078	– 7.1	10550	–107.1
28	988	– 97.1	9750	144.4	17	1085	– 0.1	10800	61.2
1	998	– 87.1	9650	– 72.4	5	1090	4.9	10650	–147.2
16	998	– 87.1	9850	+127.6	22	1130	44.9	11000	–264.5
21	1011	– 74.1	9800	– 74.3	26	1149	63.9	11400	– 86.5
9	1012	– 73.1	10100	214.0	29	1169	83.9	11900	179.8
8	1025	– 60.1	10000	– 37.9	30	1180	94.9	11750	– 98.7
12	1035	– 50.1	10300	145.3	4	1181	95.9	11600	–260.4
23	1042	– 43.1	10000	–236.5	13	1190	104.9	11900	– 65.5
25	1043	– 42.1	10200	– 48.2	19	1215	129.9	11950	–307.6
10	1047	– 38.1	10300	+ 5.1	24	1217	131.9	12200	– 80.9
20	1055	– 30.1	10500	111.6	14	1240	154.9	12350	–199.6
15	1058	– 27.1	10300	–123.1	6	1271	185.9	13800	888.2

Units are omitted in the calculations.
Mean of $x' = a = 1085.10$

tained by means of a statistical sampling *procedure,* there is a chance of making an incorrect decision. The probability of making an incorrect decision usually can be reduced by increasing the number of samples tested. Two types of wrong decisions are possible:

1. It is concluded that there is a difference but, in fact, there is none (Error of the First Kind). The probability of making this error is denoted by α.

2. It is concluded that there is no difference but, in fact, there is one (Error of the Second Kind). The probability of making this error is denoted by β.

The probability of an Error of the Second Kind is related to the size of the difference δ being measured. The value of β associated with a particular δ decreases as δ increases. For a specific statistical test, the ability to detect a difference is determined by α, β (for a given δ), and N the sample size. The quantity $1 - \beta$ (δ) is called the Power of the test to detect a difference δ with a sample of size N when the test is performed at an α level of s-significance. The relationship between these parameters can be described graphically by an Operating Characteristic (OC) curve. These curves describe the discriminatory power of a test. There is a unique OC curve that corresponds to specified values of N and α. These and other kinds of OC curves are discussed in greater detail in Chapter **3**.

Two basic types of tests can be considered:

1. Does the mean differ from a specified requirement?

2. Does the mean of one design differ from the mean of another design?

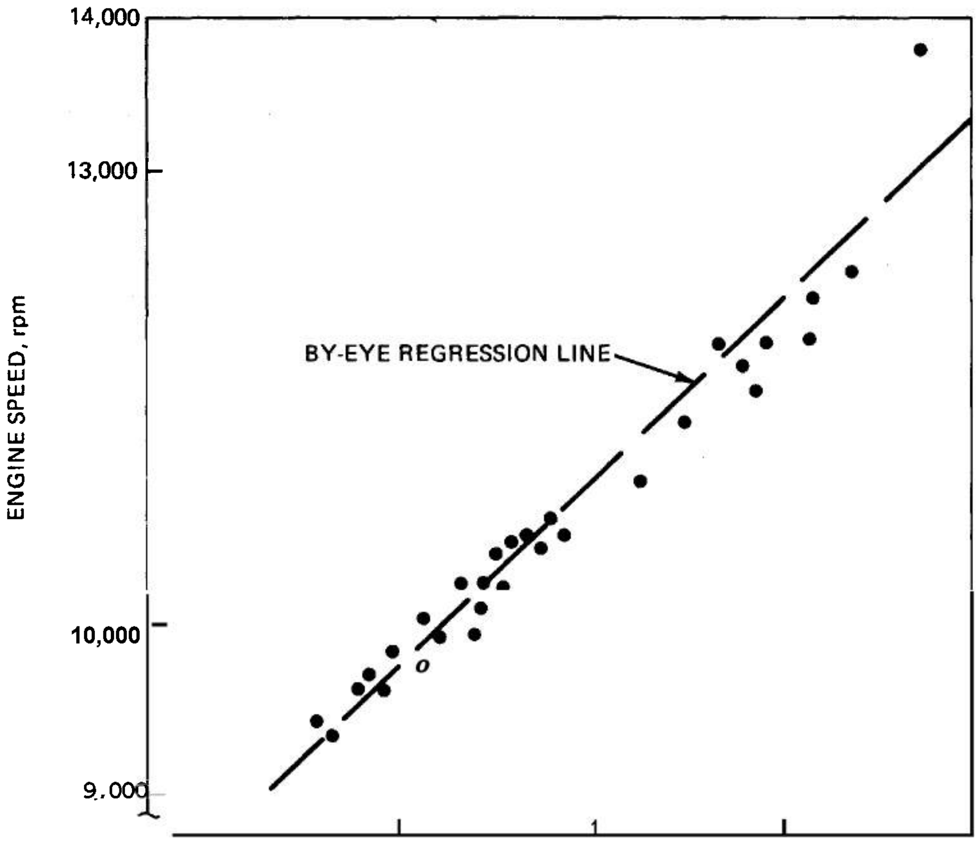


Figure 2-13. Scattergram of Test Data for Turbine Blades[2]

The first test can sometimes be used to determine if a product meets its contractual reliability requirements. The second can be used to compare one design with another (Ref. 18).

The relationships to be used in performing these tests are summarized in Tables 2-23 and 2-24. Table 2-23 presents the techniques for comparing the mean life of a product with a previously defined standard. The variance may either be known in advance or estimated from the data. Table 2-24 presents the techniques for testing one design against another. Again, the conditions of both known and unknown variance are considered. Sample sizes can be estimated from Table 2-25 and 2-26.

Example No. 15 illustrates the procedure.

2-10 ACCEPT/REJECT TESTS–BINOMIAL PARAMETER

Tests of hypothesis and s-significance can be performed on the binomial parameter. These techniques can be used to compare s-reliabilities of various designs and for comparing achieved s-reliability with a requirement. These procedures are described in Ref. 18. *[Text continues on page 2-109.]*

Example No. 15

A certain fuse is experiencing reliability problems. It is redesigned and both fuse designs are tested to determine if the new designs have *s*-significantly better reliability. The fuses of original design have an estimated mean life of 30,000 hr with an estimated standard deviation of 500 hr (100 fuses tested), and the redesigned fuses have an estimated mean life of 31,000 hr with an estimated standard deviation of 560 hr (100 fuses tested). Does the redesigned fuse have better reliability at the 90% level of *s*-confidence.

Procedure	Example
1A. State the parameters of the problem. Assume that the standard deviation estimate was the *s*-statistic.	1A. $\bar{X}_A = 31{,}000$, $\bar{X}_B = 30{,}000$ $s_A = 560$, $s_B = 500$. The units of hours will be implied for $\bar{X}$ and s. $N_A = 100$, $N_B = 100$.
1B. Make the explicit assumption of s-normality for both fuse designs.	
1C. Be explicit about the measure of reliability.	1C. The measure of reliability will be the true mean-life of the fuse.
1D. Assume $\sigma_A \approx \sigma_B$.	
2. Compute the degrees of freedom from Table 2-24 $\nu = N_A + N_B - 2$.	2. $\nu = 100 + 100 - 2 = 198$.
3. Compute s_p from Table 2-24 $s_p^2 = \dfrac{(N_A - 1)s_A^2 + (N_B - 1)s_B^2}{N_A + N_B - 2}$.	3. $s_p^2 = \dfrac{(100-1)(560)^2 + (100-1)(500)^2}{100 + 100 - 2}$ $s_p = 531$.
4. Determine the critical value of t for 1-sided 90% s-confidence, from Table 2-27.	4. For $\nu = 198$ and 1-sided 90% s-confidence, $t = 1.28$.
5. Compute $u \equiv t s_p \left(\dfrac{1}{N_A} + \dfrac{1}{N_B}\right)^{1/2}$, an uncertainty in the difference of the means.	5. $u = 128(531)\left(\dfrac{1}{100} + \dfrac{1}{100}\right)^{1/2}$ $= 96$.

6. Compute $\bar{X}_A - \bar{X}_B$.	6. $\bar{X}_A - \bar{X}_B = 31{,}000 - 30{,}000 = 1000$.
7. Compare $(\bar{X}_A - \bar{X}_B)$ to u.	7. $1000 > 96$.

The redesigned fuse has a greater mean life than the original one at 90% level of s-confidence. The difference is s-significant at the 10% level,: indeed, it is s-significant at the 0.5% level (and probably at any feasible level). But is it significant, i. e., is it important in an engineering sense? The improvement is small, approximately 3% in mean life. There might be a slight degradation in standard deviation, although not an important one. If the new design has any disadvantages at all, it may be better to leave things as they are; 3% is a small change, it could be lost, for example, in month-to-month or batch-to-batch variations.

TABLE 2-23

SUMMARY OF TECHNIQUES FOR COMPARING THE AVERAGE OF A NEW PRODUCT WITH THAT OF A STANDARD[18]

We Wish To Test Whether	Knowledge of Variation of New Item	Test To Be Made	Sample Size Required	Notes
m differs from m_0	σ unknown; s = estimate of σ from sample.	$\lvert \bar{X} - m_0 \rvert > u$	Use Table **2-25** for α = **0.05**, add **2** to tabular value. For α = 0.01, add **4** to tabular value.	$u = t_{1-\alpha/2}\left(\frac{s}{\sqrt{N}}\right)$ (t for $N - 1$ degrees of freedom
	σ known	$\lvert \bar{X} - m_0 \rvert > u$	Use Table **2-25**	$u = z_{1-\alpha/2}\left(\frac{\sigma}{\sqrt{N}}\right)$
m is larger than m_0	σ unknown; s = estimate of σ from sample	$(\bar{X} - m_0) > u$	Use Table **2-26** for α = 0.05, add **2** to tabular value. For α = 0.01, add **3** to tabular value.	$u = t_{1-\alpha}\left(\frac{s}{\sqrt{N}}\right)$ (t for $N - 1$ degrees at freedom)
	σ known	$(\bar{X} - m_0) > u$	Use Table **2-26**	$u = z_{1-\alpha}\left(\frac{\sigma}{\sqrt{N}}\right)$
m is smaller than m_0	σ unknown; s = estimate of σ from sample.	$(m_0 - \bar{X}) > u$	Use Table **2-26** for α = **0.05**, add **2** to tabular value. For α = 0.01, add **3** to tabular value.	$u = t_{1-\alpha}\left(\frac{s}{\sqrt{N}}\right)$ (t for $N - 1$ degrees of freedom)
	σ known	$(m_0 - \bar{X}) > u$	Use Table **2-26**	$u = z_{1-\alpha}\left(\frac{\sigma}{\sqrt{N}}\right)$

t = standard Student t variate; see Table **2-27.**
z = standard s-normal variate
m = mean life at new product
m_0 = mean life of standard
$\bar{X}$ = arithmetic mean of the new-product test data.

TABLE 2-24

SUMMARY OF TECHNIQUES FOR COMPARING THE AVERAGE PERFORMANCE OF TWO PRODUCTS[18]

We Wish To Test Whether	Knowledge of Variation	Tests to be Made	Determination of Sample Size N	Notes
m_A differs from m_B	$\sigma_A \approx \sigma_B$; both unknown	$\lvert \bar{X}_A - \bar{X}_B \rvert > u$, where $u = (t_{1-\alpha/2}) s_p \sqrt{\dfrac{N_A + N_B}{N_A N_B}}$	Use Table 2-25. For $\alpha = 0.05$, add 1 to tabular value. For $\alpha = 0.01$, add 2 to tabular value.	$s_p = \sqrt{\dfrac{(N_A - 1)s_A^2 + (N_B - 1)s_B^2}{\nu}}$ $\nu = N_A + N_B - 2$
	$\sigma_A \neq \sigma_B$; both unknown	$\lvert \bar{X}_A - \bar{X}_B \rvert > u$, where $u = t' \sqrt{\dfrac{s_A^2}{N_A} + \dfrac{s_B^2}{N_B}}$		t is the value of $t_{1-\alpha/2}$ for the effective number of degrees of freedom $\nu = \dfrac{(s_A^2/N_A + s_B^2/N_B)^2}{\dfrac{(s_A^2/N_A)^2}{N_A + 1} + \dfrac{(s_B^2/N_B)}{N_B + 1}} - 2$
	σ_A, σ_B both unknown	$\lvert \bar{X}_A - \bar{X}_B \rvert > u$, where $u = z_{1-\alpha/2} \sqrt{\dfrac{\sigma_A^2}{N_A} + \dfrac{\sigma_B^2}{N_B}}$	Use Table 2-25	

TABLE 2-24 (Cont'd)

We Wish To Test Whether	Knowledge of Variation	Tests to be Made	Determination of Sample Size N	Notes
$m_A > m_B$	$\sigma_A \approx \sigma_B$; both unknown	$(\bar{X}_A - \bar{X}_B) > u$, where $u = (t_{1-\alpha}) s_P \sqrt{\frac{N_A + N_B}{N_A N_B}}$	Use Table **2-26** For $\alpha = 0.05$, add **1** to tabular value. For $\alpha = \mathbf{0.01}$, add **2** to tabular value.	$s_P = \sqrt{\frac{(N_A - 1)s_A^2 + (N_B - 1)s_B^2}{\nu}}$ $\nu = N_A + N_B - 2$
	$\sigma_A \neq r_B$; both unknown	$(\bar{X}_A - \bar{X}_B) > u$, where $u = t' \sqrt{\frac{s_A^2}{N_A} + \frac{s_B^2}{N_B}}$		t' is the value of $t_{1-\alpha}$ for the effective number of degrees of freedom $\nu = \frac{(s_A^2/N_A + s_B^2/N_B)^2}{\frac{(s_A^2/N_A)^2}{N_A + 1} + \frac{(s_B^2/N_B)^2}{N_B + 1}} - 2$
	σ_A, σ_B both unknown	$(\bar{X}_A - \bar{X}_B) > u$, where $u = z_{1-\alpha} \sqrt{\frac{\sigma_A^2}{N_A} + \frac{\sigma_B^2}{N_B}}$	Use Table **2-26**	

ν = degrees of freedom
t = standard Student t variate; see Table **2-27**
z = standard s-normal variate
$\underline{m}$ = mean life of product, used with subscript **A** or B
X = arithmetic mean of product test data, used with subscript **A** or B
subscript **A** = product **A**
subscript B = product B

TABLE 2-25 SAMPLE SIZES REQUIRED TO DETECT PRESCRIBED DIFFERENCES BETWEEN AVERAGES WHEN THE SIGN OF THE DIFFERENCE IS NOT IMPORTANT[18]

The table entry is the sample size (N) required to detect, with probability $1-\beta$, that the average m of a new product differs from the standard m_0 (or that two product averages m_A and m_B differ). The standardized difference is d, where

$d = \frac{|m - m_0|}{\sigma}$ (or $d = \frac{|m_A - m_B|}{\sigma}$ if we are comparing two products).

The standard deviations are assumed to be known, and N is determined by the formula:

$$N = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2}{d^2}$$

$\alpha = .01$

d \ $1-\beta$	.50	.60	.70	.80	.90	.95	.99
.1	664	801	962	1168	1488	1782	2404
.2	166	201	241	292	372	446	601
.4	42	51	61	73	93	112	151
.6	19	23	27	33	42	50	67
.8	11	13	16	19	24	28	38
1.0	7	9	10	12	15	18	25
1.2	5	6	7	9	11	13	17
1.4	4	5	5	6	8	10	13
1.6	3	4	4	5	6	7	10
1.8	3	3	3	4	5	6	8
2.0	2	3	3	3	4	5	7
3.0	1	1	2	2	2	2	3

If we must estimate σ from our sample and use Student's t, then we should add 4 to the tabulated values to obtain the approximate required sample size. (If we are comparing two product averages, add 2 to the tabulated values, to obtain the required size of each sample. For this case, we must have $\sigma_A = \sigma_B$.).

$\alpha = .05$

d \ $1-\beta$	.50	.60	.70	.80	.90	.95	.99
.1	385	490	618	785	1051	1300	1838
.2	97	123	155	197	263	325	460
.4	25	31	39	50	66	82	115
.6	11	14	18	22	30	37	52
.8	7	8	10	13	17	21	29
1.0	4	5	7	8	11	13	19
1.2	3	4	5	6	8	10	13
1.4	2	3	4	5	6	7	10
1.6	2	2	3	4	5	6	8
1.8	2	2	2	3	4	5	6
2.0	1	2	2	2	3	4	5
3.0	1	1	1	1	2	2	3

If we must estimate σ from our sample and use Student t, then we should add 2 to the tabulated values to obtain the approximate required sample size. (If we are comparing two product averages, add 1 to the tabulated values to obtain the required size of each sample. For this case, we must have $\sigma_A = \sigma_B$.).

TABLE 2-26 SAMPLE SIZES REQUIRED TO DETECT PRESCRIBED DIFFERENCES BETWEEN AVERAGES WHEN THE SIGN OF THE DIFFERENCE IS IMPORTANT[18]

The table entry is the sample size (N) required to detect with probability $1 - \beta$ that:
(a) the average m of a new product exceeds that of a standard m_0
(b) the average m of a new product is less than that of a standard m_0
(c) the average of a specified product m_A exceeds the average of another specified product m_B.
The standardized difference is d, where:

$$\text{(a)}\quad d = \frac{m - m_0}{\sigma}$$

$$\text{(b)}\quad d = \frac{m_0 - m}{\sigma}$$

$$\text{(c)}\quad d = \frac{m_A - m_B}{\sqrt{\sigma_A^2 + \sigma_B^2}}$$

The standard deviations are assumed to be known, and N is calculated from the following formula:

$$N = \frac{(z_{1-\alpha} + z_{1-\beta})^2}{d^2}$$

$\alpha = .01$

d \ $1-\beta$	.50	.60	.70	.80	.90	.95	.99
.1	542	666	813	1004	1302	1578	2165
.2	136	167	204	251	326	395	542
.4	34	42	51	63	82	99	136
.6	16	19	23	28	37	44	61
.8	9	11	13	16	21	25	34
1.0	6	7	9	11	14	16	22
1.2	4	5	6	7	10	11	16
1.4	3	4	5	6	7	9	12
1.6	3	3	4	4	6	7	9
1.8	2	3	3	4	5	5	7
2.0	2	2	3	3	4	4	6
3.0	1	1	1	2	2	2	3

If we must estimate σ from our sample, and use Student's t, add 3 to the tabulated values to obtain the approximate required sample size. (If we are comparing two product averages, add 2 to the tabulated values to obtain the required size of each sample. For this case, we must have $\sigma_A = \sigma_B$).

$\alpha = .05$

d \ $1-\beta$	.50	.60	.70	.80	.90	.95	.99
.1	271	361	471	619	857	1083	1578
.2	68	91	118	155	215	271	395
.4	17	23	30	39	54	68	99
.6	8	11	14	18	24	31	44
.8	5	6	8	10	14	17	25
1.0	3	4	5	7	9	11	16
1.2	2	3	4	5	6	8	11
1.4	2	2	3	4	5	6	9
1.6	2	2	2	3	4	5	7
1.8	1	2	2	2	3	4	5
2.0	1	1	2	2	3	3	4
3.0	1	1	1	1	1	2	2

If we must estimate σ from our sample, and use Student t, add 2 to the tabulated values to obtain the approximate required sample size. (If we are comparing two product averages, add 1 to the tabulated values to obtain the required size of each sample. For this case, we must have $\sigma_A = \sigma_B$).

TABLE 2-27

PERCENTILES OF THE STUDENT t-DISTRIBUTION[18]

The body of the table gives the Cdf.

ν	$t_{.60}$	$t_{.70}$	$t_{.80}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	.289	.617	1.061	1.886	2.920	4.303	6.965	9.925
3	.277	.584	0.978	1.638	2.353	3.182	4.541	5.841
4	.271	.569	.941	1.533	2.132	2.776	3.747	4.604
5	.267	.559	.920	1.476	2.015	2.571	3.365	4.032
6	.265	.553	.906	1.440	1.943	2.447	3.143	3.707
7	.263	.549	.896	1.415	1.895	2.365	2.998	3.499
8	.262	.546	.889	1.397	1.860	2.306	2.896	3.355
9	.261	.543	.883	1.383	1.833	2.262	2.821	3.250
10	.260	.542	.879	1.372	1.812	2.228	2.764	3.163
11	.260	.540	.876	1.363	1.796	2.201	2.718	3.106
12	.259	.539	.873	1.356	1.782	2.179	2.681	3.055
13	.259	.538	.870	1.350	1.771	2.160	2.650	3.012
14	.258	.537	.868	1.345	1.761	2.145	2.624	2.977
15	.258	.536	.866	1.341	1.753	2.131	2.602	2.947
16	.258	.535	.865	1.337	1.746	2.120	2.583	2.921
17	.257	.534	.863	1.333	1.740	2.110	2.567	2.898
18	.257	.534	.862	1.330	1.734	2.101	2.552	2.878
19	.257	.533	.861	1.328	1.729	2.093	2.539	2.861
20	.257	.533	.860	1.325	1.725	2.086	2.528	2.845
21	.257	.532	.859	1.323	1.721	2.080	2.518	2.831
22	.256	.532	.858	1.321	1.717	2.074	2.508	2.819
23	.256	.532	.858	1.319	1.714	2.069	2.500	2.807
24	.256	.531	.857	1.318	1.711	2.064	2.492	2.797
25	.256	.531	.856	1.316	1.708	2.060	2.485	2.787
26	.256	.531	.856	1.315	1.706	2.056	2.479	2.779
27	.256	.531	.855	1.314	1.703	2.052	2.473	2.771
28	.256	.530	.855	1.313	1.701	2.048	2.467	2.763
29	.256	.530	.854	1.311	1.699	2.045	2.462	2.756
30	.256	.530	.854	1.310	1.697	2.042	2.457	2.750
40	.255	.529	.851	1.303	1.684	2.021	2.423	2.704
60	.254	.527	.848	1.296	1.671	2.000	2.390	2.660
120	.254	.526	.845	1.289	1.658	1.980	2.358	2.617
∞	.253	.524	.842	1.282	1.645	1.960	2.326	2.576

The table in Ref. 18 was extracted from a larger one in Ref. 21.

2-11 ACCEPT/REJECT TESTS-NONPARAMETRIC

Nonparametric tests can be used to evaluate many properties of a distribution or to make comparisons, Ref. 25. The tests in this paragraph deal with the following kind of experiment (Refs. 16 and 25). A sample of N items is life tested in the usual (u) environment. Times to failure are noted. A similar set of N items is simultaneously subjected to a more severe (s) environment. The statistical analysis determines if exposure to "s" changes the life of the units in a s-significant sense.

The null hypothesis (no difference in the quality that the statistical test measures) is tested against the alternative that there is a difference. Engineers need not concern themselves about the technical details of such hypotheses.

Three test procedures will be described in these paragraphs: (1) rank-sum, (2) run, (3) maximum-deviation–Example Nos. 16, 17, 18, and 19. They will be described for the same basic set of data. The successful application of these tests depends on the fact that life-test data can be put in time order.

2-11.1 RANK-SUM

Also known as the Mann-Whitney or Wilcoxon test. See Sec. 5.3, Ref. 25, and Example No. 17.

2-11.2 RUNS

Also known as the Wald-Wolfowitz test. See Ref. 16 or Sec. 7.3 of Ref. 25, and Example No. 18.

If the smaller of the rank sums is greater than the critical number in the body of the table, accept the null hypothesis that the chances are 50%–50% that any unit from one population will have a shorter life than any unit from the other population.

Both the rank-sum test and the run test require waiting for all items in both samples to fail. The next test permits test truncation and a shorter test time.

2-11.3 MAXIMUM-DEVIATION

The maximum-deviation test is a truncated test (Ref. 16). In this test, a value is preassigned to r, where r refers to the rth order statistic. For example, $r = 2$. Then examine the data to establish the time (either u_2 or s_2) at which at least two failures have occurred in both samples. From the test data, the time corresponding to $r = 2$ is $u_2 = 7.5$, at which time two failures have occurred in the "usual" sample and three failures have occurred in the "severe" sample.

Define a quantity M_r as the absolute difference between the number of failures in the u and s samples calculated after each failure. Two other required parameters, m_r and p_r, are tabulated in Table 2-32 for r = 1, 3, 6, 10, and for samples of size 10. (An expanded version of this table is available in Ref. 16.) For any r, $t_r = \max$ u_r, s_r .

Proceed as follows. If $M_r = m_r + 1$ at any time up to and including t_r, stop the test and reject H_0. If $M_r \leqslant (m_r - 1)$ up to and including t_r, accept H_0. If the test is continued to t_r, and $M_r = m_r$ at least once and otherwise $M_r < (m_r - 1)$, perform a Bernoulli trial which will reject H_0 with probability p_r. See Example No. 19.

2-12 SYSTEM RELIABILITY ESTIMATION FROM SUBSYSTEM DATA

Many weapon systems are extremely complex and consist of large numbers of subsystems and components. During a considerable portion of the system development cycle, only component and subsystem failure data are available for reliability analyses. The amount of data often varies considerably

[text continues on page 2-114.]

Example No. 16 (Ref. 14)

The fire control subsystem of a tank is tested to determine if a high vibration environment influences the life. A "usual" sample and "severe" sample of 10 times-to-failure each are obtained. The ordered times-to-failure are recorded in hours; see Table 2-28.

TABLE 2-28

LIFE DATA, FIRE CONTROL SYSTEM

(A) Usual environment (N = 10)		(B) Severe environment (N = 10)	
4.0	12.8	1.0	9.5
7.5	13.0	3.5	15.0
8.0	14.0	6.5	19.5
9.0	21.5	8.5	28.0
11.5	27.5	8.9	31.1

These data can be combined into an ordered array of 20, but with their identity noted (see Table 2-29).

TABLE 2-29

COMBINED, IDENTIFIED DATA

1. os	8. 0u	11.5u	19.55
3.5s	8.5s	12.8u	21.5u
4.0u	8.9s	13. Ou	27.5u
6.5s	9. 0u	14. Ou	28. Os
7.5u	9.5s	15. Os	31.15

Example No. 17

Procedure	Example
1. Assign the ranks 1, 2, . . ., 20 to the ordered sample of twenty, i. e., the sample 1.0 is labelled 1 and the sample 31.1 is labelled 20.	1. (see table below)
2. Identify the ranks of sample u.	2. 3, 5, 6, 9, 11, 12, 13, 14, 17, 18.
3. Identify the ranks of sample s.	3. 1, 2, 4, 7, 8, 10, 15, 16, 19, 20.
4. Compute the sum of the ranks of sample u.	4. $\Sigma \text{rank}_u = 108$.
5. Compute the sum of the ranks of sample s.	5. $\Sigma \text{rank}_s = 102$.
6. From Table 2-30, find the critical rank-sum s-significance number C.	6. From Table 2-30, for an original sample size N = 10 and s-significance level = 5%. $C = 79$.
7. Test the null hypothesis.	7. Since the smaller rank sum, $\Sigma \text{rank}_{us} = 102$, is larger than $C = 79$, accept the hypothesis stated in Table 2-30.

Sample	Rank	Sample	Rank
1.0s	1	11.5u	11
3.5s	2	12.8u	12
4.0u	3	13.0u	13
6.5s	4	14.0u	14
7.5u	5	15.0s	15
8.0u	6	19.5s	16
8.5s	7	21.5u	17
8.9s	8	27.5u	18
9.0u	9	28.0s	19
9.5s	10	31.1s	20

Thus, on the basis of this test, we presume that a random item from one population is equally likely to last longer or shorter than a random item from the other population.

Example No. 18

Procedure	Example
1. Tabulate the full ordered sample of 20 and mark each item with a u if it came from the u sample and an s if it came from the s sample.	1. See Table 2–29.
2. A succession of u's or s's is called a run (a single u or s is a run of one). Count the runs.	2. Σ runs = 11.
3. Enter Table 2–31 to determine an acceptance number A and a rejection number R for s-significance level = 5%.	3. From Table 2 3 1 for $N = 10$, $A = 8$ $R = 6$.
4. Determine the validity of the hypothesis.	4. Since Σ runs = 11, and $11 > 8$, accept the null hypothesis that the two Cdf's are the same.

Example No. 19

Procedure	Example
1. For $r = 1$, determine u_1 and s_1.	1. $u_1 = 4.0$ $s_1 = 1.0$.
2. Find $t_1 \equiv \max\{u_1, s_1\}$.	2. $t_1 = 4.0$.
3. Define M_r as the absolute difference between the number of failures in the u and s samples computed after each failure occurs. Compute this M_r until the first failure occurs in both u and s.	3. For $r = 1$, after first failure, $M_r = 1$ after second failure, $M_r = 2$ after third failure, $M_r = 1$. Stop here because failure in u = failure in s.
4. Determine m_r from Table 2-32. s-significance level = 5%.	4. $r = 1$, $m_1 = 4$.
5. Evaluate the inequality. If $M_r < (m_r - 1)$, accept H_0.	5. $1 < (4 - 1) = 3$. Therefore we accept the hypothesis H_0 at the 5% s-significance level.
6. Repeat for $r = 3$. For $r = 3$, determine u_r and s_r.	6. $u_3 = 8.0$ $s_3 = 6.5$.
7. Find $t_3 \equiv \max\{u_3, s_3\}$	7. $t_3 = 8.0$.
8. Compute M_r until the third failure occurs in both u and s.	8. For $r = 3$, after first failure, $M_r = 1$ after second failure, $M_r = 2$ after third failure, $M_r = 1$ after fourth failure, $M_r = 2$ after fifth failure, $M_r = 1$ after sixth failure, $M_r = 0$.
9. Determine m_r from Table 2-32 at 5% s-significance level.	9. $r = 3$ $m_3 = 5$.
10. Evaluate the inequality. If $M_r < (m_r - 1)$, accept H_0.	10. $3 < (5 - 1) = 4$ Therefore we accept the hypothesis H_0 at the 5% level of s-significance.

TABLE 2-30*

RANK-SUM TEST
s-SIGNIFICANCE CRITERIA'[6]

s-Significance level

N	0.05	0.02	0.01
5	18	16	15
6	27	24	23
7	37	34	32
8	49	46	43
9	63	59	56
10	79	74	71
11	97	91	87
12	116	110	105
13	137	130	125
14	160	152	147
15	185	176	170
16	212	202	196
17	241	230	223
18	271	259	252
19	303	291	282
20	338	324	315

N = number of items in each sample.

TABLE 2-31*

RUN-TEST s-SIGNIFICANCE CRITERIA'[6]

s-Significance Level = 5%

n	R (reject H_0 if $\Sigma_{runs} \leqslant R$)	A (accept H_0 if $\Sigma_{runs} \geqslant A$)	D (if $\Sigma_{runs} = D$) issue is in doubt)
5	3	4	
6	3	4	
7	4	5	
8	5	6	
9	6	7	
10	6	8	7 (H_0 disfavored)
11	7	9	8 (H_0 favored)
12	8	10	9 (H_0 favored)
13	9	11	10 (H_0 favored)
14	9	11	10 (H_0 disfavored)
15	10	12	11 (H_0 disfavored)
16	11	13	12 (H_0 slightly favored)
20	14	16	15 (H_0 disfavored)

H_0 = null hypothesis = there is no difference in the *Cdf's* of the 2 variables.

*Reprinted from *Mathematical Methods in Reliability Engineering* with permission of McGraw-Hill Book Company, Inc.

from subsystem to subsystem, depending on their availability for testing and the amount of previous experience with similar items. Techniques have been developed for estimating system reliability based on subsystem, equipment, and component reliability data.

Of course, estimating system reliability from lower level data can never be as accurate as direct testing of the assembled system. However, these estimates are very useful for decision making before completed systems are available for testing.

2-12.1 ADVANTAGES OF MODEL

This paragraph describes a statistical model that can be used by contractors for reliability estimation and by the Army for weapon-system reliability-monitoring. The model permits the combination of test data from all levels (from component to weapon system) and all types of tests into meaningful component, equipment, and subsystem failure rates and reliability predictions. The failure rates and reliability estimates can be updated continuously as new test data become available.

TABLE 2-32

MAXIMUM-DEVIATION-TEST s-SIGNIFICANCE CRITERIA'[6]

5% s-Significance Level

Sample size: N = 10

r	m_r	p_r
1	4	0.32
3	5	0.17
6	6	0.95
10	6	0.95

The information required for the use of this statistical model is listed here:

1. Detailed analysis of the mission profile to determine stress levels and durations for each hardware level and to convert actual test time to mission equivalents.

2. The subsystem (or system) reliability equation.

3. Test data, including total test times and associated environmental stresses and failures from all test sources.

The model permits reliability assessment to begin on a piecemeal basis with information derived from development tests, engineering evaluation tests, and from qualification tests. Later on, information from production tests also can be used.

The tests may differ with respect to levels, conditions, and durations; they may be applied at component, equipment, and subsystem levels, and the test conditions include a wide variety of environmental conditions; and their durations may differ considerably.

The calculation of the best estimate of component failure rate in a specific environment will be discussed first. Then techniques for incorporating data from higher level tests (equipment and subsystem) in the calculation of component failure rates will be presented. Once the basic failure rate is established, methods are developed for combining these estimates to generate the best estimate of mission reliability. A procedure for estimating the uncertainty of these estimates also is described.

2-12.2 COMPONENT MODEL

Consider components in a complex system which must operate successfully over a defined mission (Ref. 33). During the mission, the Component is exposed to environmental stresses of different degrees, kinds, and durations. The components must perform their assigned functions when they are needed. The probability of completing these functions satisfactorily in the operating environment is called s-reliability. The total time that a specific component must operate may be less than the total mission time. Reliability is, therefore, a function of the environmental and usage stresses, the operating and nonoperating conditions, the part reference-failure rates, and the time duration of the environments and usage stresses. The component reliability model is based on the following assumptions (modified from Ref. 33):

1. Failure rate is independent of time.

2. Part failures are s-independent of each other.

3. Part failure rate is independent of the history of the component.

4. A specific assumption must be made about the failure rate for all phases of the mission, even benign subphases.

5. There is a great deal of uncertainty in the estimation process, much of which is related to the suitability of the model, and the remainder is due to statistical uncertainty (calculated presuming the model is perfect).

6. The k-factor approach is useful for converting failure rates from one set of conditions to another, especially for storage-like conditions.

7. Each special environment can be represented by an additive term onto the reference failure rate. The size of the term is *s*-independent of the presence of other terms. It is as if each special environment aggravated exactly one sindependent failure mode in the item.

Due to Assumption No. 5 there is little point in making minor statistical refinements in the estimate. For example, s-bias in the estimate of failure rate is of little concern, especially since lack-of-s-bias is preserved only in linear transformations. If $\hat{\theta}$ is an s-unbiased estimate of "$\theta = 1/\lambda$", then $1/\hat{\theta}$ is a s-biased estimate of λ and R exp $(-t/\hat{\theta})$ is a s-biased estimate of survival probability. Since no one knows which function "ought" to be s-unbiased, s-bias is ignored.

The equation for estimating the failure rate is the maximum likelihood equation

$$\hat{\lambda} = r/T \tag{2-67}$$

where

$\hat{\lambda}$ = estimated failure rate

r = total failures

T = total test time

If the mission is partitioned into phases and subphases, an equivalent failure rate for the mission can be developed, if desired, by a simple averaging process:

$$\lambda_{equiv} = \Sigma_\alpha \lambda_\alpha \phi_\alpha \tag{2-68}$$

where

α = mission phase or subphase

Σ_α = implies sum over all phases (subphases) of a partitioned mission

ϕ_α = fraction of mission time spent in α

λ_α = failure rate while in α

The coefficient of variation of $\hat{\lambda}$ is (for large r)

$$\eta_{\hat{\lambda}} = 1/\sqrt{r+1} \tag{2-69}$$

For small r, the formula is not accurate and there is no simple rigorous formula. Eq. 2-69 is well within the limitations of Assumption No. 5. The choice of modification for small r depends on lots of things; it is hard to fault any approach with even a modicum of reason to it. The +1 with the r merely keeps the equation from "blowing up" at $r = 0$.

In complicated equations, it is usually sufficient to estimate the standard-deviation or coefficient-of-variation of a function. If s-confidence limits are desired, when using Eq. 2-67, there are no exact ways to get them because (a) there is a lack of knowledge about the details of the tests and (b) the method of combining data causes difficulties. See pars. **2-3.2** and **2-3.3** for details.

Example Nos. 20 and 21 illustrate this statistical model.

2-12.3 SYSTEM MODEL

For series systems, the system failure rate is a linear function of the component failure rates. For linear functions with s-independent variables, the mean of the function is the function of the means and the variance of the function is the function of the variances. For more complicated systems, see ***Part Two, Design for Reliability*** the paragraph on Parameter Variations Analysis.

For further discussion of this and similar models, see Refs. **33-35,** but treat any procedure with skepticism that seems to violate Assumption No. 5 (uncertainty) or pushes Assumption No. 7 (additive failure rates for environments) too far.

Example No. 20

Three components are tested: No. 1 was censored after 3 hr; No. 2 was censored after 6 hr; No. 3 failed in 1 hr; i. e., $r = 1$. The parts undergo the test conditions during 15% of the mission and are in benign circumstances otherwise. Determine the equivalent failure rate for the mission.

Procedure	Example
1. State the failure rate assumption for all phases of mission.	1. During operating phase, the failure rate is the same as on the test. During benign phase, it is assumed to be 5% of the "operating" value.
2. Estimate the "operating" failure rate. Use Eq. 2-67.	2. $\hat{\lambda}_{op} = 1/(3 + 6 + 1)\text{-hr}$ $= 0.1$ per hr.
3. Estimate the "benign" failure rate.	3. $\hat{\lambda}_{ben} = 5\%\hat{\lambda}_{op}$ $= 0.005$ per hr.
4. State the fraction of time in each phase.	4. $\phi_{op} \equiv 0.15$ $\phi_{ben} = 0.85$.
5. Calculate the equivalent mission failure rate. Use Eq. 2-68.	5. $\lambda_{eq} \times \text{hours} = 0.1 \times 0.15 + 0.005 \times 0.85 = 0.019$.
6. Estimate the coefficient of variation for $\hat{\lambda}_{op}$. Use Eq. 2-69.	6. $\eta_{\hat{\lambda}_{op}} = 1/\sqrt{1+1} = 0.71$.

The estimates of failure rates for a given test condition can be updated as more test data become available from component level testing. Tests conducted at higher levels (such as subsystem or equipment tests) which are monitored for failure causes also can provide data for updating failure rates. These data must indicate the actual time that each component operates in each mission phase (subphase); see Example No. 21.

Example No. 21

One type of component in Example No. 20 is operated in a subsystem test for 1 hr without failure. The test conditions are the same as in Example No. 20, and the component operates all the time the subsystem operates. Reevaluate the equivalent failure rate for the mission.

Procedure	Example
1. State the old cumulative test-time and failures for the "operating" condition.	1. $r_{old} = 1$, $T_{old} = 10$ hr.
2. State the present increments.	2. $\Delta r = 0$, $\Delta T = 1$.
3. Calculate the new, r, T and $\hat{\lambda}_{op}$.	3. $r_{new} = 1 + 0 = 1$, $T_{new} = (10 + 1)$hr $= 11$ hr $\hat{\lambda}_{op} = 1/11$ per hr $= 0.091$ per hr.
4. Estimate the coefficient of variation for $\hat{\lambda}$. See Example No. 20, step 6.	4. Since $r_{new} = r_{old}$, $\eta_{\hat{\lambda}}$ is the same as in Example No. 20.
5. Calculate the equivalent mission failure rate. See Example No. 20, step 5.	5. $\lambda_{eq} \times$ hours $= 0.091 \times 0.15$ $+ 0.091 \times 0.05 \times 0.85 = 0.018$.

REFERENCES

1. W. Grant Ireson, *Reliability Handbook,* McGraw-Hill Book Company, New York, NY, 1966.

2. *Handbook, Reliability Engineering,* NAVWEPS 00-65-502, Bureau of Naval Weapons.

3. *Reliability and Maintainability Handbook for the US Weather Bureau,* Publication 530-01-1-762, ARINC Research Corporation, Annapolis, MD, April 1967.

4. W. Yurkowsky, *Nonelectronic Notebook,* RADC-TR-69-458, Hughes Aircraft Company, Prepared for Rome Air Development Center, March 1970.

5. W. H. Von Alven, Ed., *Reliability Engineering,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1964.

6. S. Ehrenfeld and S. B. L. Mauer, *Introduction to Statistical Method,* McGraw-Hill Book Company, New York, NY, 1964.

7. M. Klerer and G. A. Korn, *Digital Computer User's Handbook,* McGraw-Hill Book Company, New York, NY, 1967.

8. R. Burington and D. May, *Handbook of Probability and Statistics with Tables,* Handbook Publishers, Inc., Sandusky, OH, 1952.

9. D. Owen, *Handbook of Statistical Tables,* Addison-Wesley Publishing Company, Inc., Reading, MA, 1962.

10. I. Bazovsky, *Reliability Theory and Practice,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1961.

11. *Reliability and Maintainability Status Report, Inertial Navigational Set AN/ASn-86,* US Army Electronics Command, Product Assurance Directorate, December 1970.

12. A. Hald, *Statistical Theory with Engineering Applications,* John Wiley & Sons, Inc., New York, NY, 1952.

13. A. Hald, *Statistical Tables and Formulas,* John Wiley & Sons, Inc., New York, NY, 1952.

14. *M60A1E2 Tank, Reliability Status Report,* Defense Operations Division, Chrysler Corporation, Prepared for Army Materiel Command, Project Manager–M60 and M60A1 Tanks, October 1970.

15. M. V. Johns and G. J. Lieberman, "An Exact Asymptotically Efficient Confidence Bound for Reliability in the Case of the Weibull Distribution", *Technometrics* **8,** No. 1, (February 1966).

16. N. H. Roberts, *Mathematical Methods in Reliability Engineering,* McGraw-Hill Book Company, New York, NY, 1964.

17. A. H. Bowker and G. J. Lieberman, *Handbook of Industrial Statistics,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1955.

18. Mary G. Natrella, *Experimental Statistics,* National Bureau of Standards Handbook 91, October 1966. Also, AMCP 706-110 through -114, Engineering Design Handbooks.

19. S. Siegel and J. W. Tukey, "A Nonparametric Sum of Ranks Procedure for Relative Spread in Unpaired Samples", *Joumal of the American Statistical Association* **55,** 429 (1960).

20. F. S. Swed and C. Eisenhart, "Tables for Testing the Randomness of Grouping in a Sequence of Alternatives",

Annals of Mathematical Statistics **14,** 66-87 (1943).

21. W. J. Dixon and **F. J.** Massey, Jr., *Introduction to Statistical Analysis,* McGraw-Hill **Book** Company, New York, NY, 1957.

22. G. **J.** Hahn and **W.** Nelson, "A Comparison **of** Methods for Analyzing Censored Life Data to Estimate Relationships Between Stress and Product Life", *IEEE Transactions on Reliability,* Vol. R-23, April 1974 (contains many references to other papers).

23. W. Nelson, "Theory and Application of Hazard Plotting for Censored Failure Data", *Technometrics* **14,** 945-966 (1972).

24. Kendall and Stuart, "Inference and Relationship", Vol. 2, *The Advanced Theory of Statistics,* 3rd ed., 1973, Hafner Publ. Co., NY.

25. W. **J.** Conover, *Practical Nonparametric Statistics,* 1971, John Wiley & Sons, NY.

26. Anon., "Tables of Plotting Positions", Vol. 3, *Failure Distribution Analyses Study,* Computer Applications, Inc., New York, NY, 10 Aug 1964, AD-631 527.

27. R. **A.** Evans, "Parameter Variations Analysis", *Practical Reliability,* **Vol.** I, November 1967, N68-31478; NASA CR-1126, July 1968.

28. W. Nelson, "Hazard Plotting for Incomplete Failure Data", J. of Quality Technology **1,** 27-52 (1969).

29. W. Nelson, "Hazard Plotting Methods for Analysis of Life Data with Different Failure Modes", **J.** of Quality Technology 2, 126-149 (1970).

30. Pearson and Hartley, *Biometrika Tables for Statisticians,* Vol. I, Cambridge Univ. Press, 1956.

31. P. I. Feder, "Some Differences Between Fixed, Mixed, and Random Effects Analysis of Variance Models", J. Quality Technology ***6,*** 98-106 (1974).

32. A. **J.** Duncan, *Quality Control and Industrial Statistics,* 3rd ed., R. D. Irwin, Inc., Homewood, Il., 1965.

33. *Guide Manual **for** Reliability Measurement Program,* NAVWEPS OD 29304, Dept. of the Navy, Special Projects Office, 15 May 1965.

34. **J.** R. Johnson, *Confidence Interval Estimation* of *the Reliability of Multicomponent Systems, using Component Test Data,* TR 3, Aberdeen Research and Development Center, November 1969.

35. D. **S.** Orkand, *A Monte Carlo Method **for** Determining Lower Confidence Limits **for** System Reliability on the Basis* of *Sample Component Data,* Report No. ORDBB-VC-4, Picatinny Research Center, Dover, **NJ,** June 1960.

CHAPTER 3

STATISTICAL EVALUATION OF RELIABILITY TESTS, DEMONSTRATIONS, AND ACCEPTANCE

LIST OF SYMBOLS

a, b = parameters of accept and reject decision lines, see Eq. 3-11

a_H, a_L = See Eq. 3-36

A, R = subscripts: A implies Accept, R implies Reject

ARL = acceptable reliability level

ASN = average sample number

c = acceptance number

c_i = acceptance number for interval i in m-sample plans

$csqf\,(\chi^2\,;\,\nu)$ = *Cdf* of the chi-square distribution with ν degrees-of-freedom

$csqfc\,(\chi^2\,;\,\nu)$ = complement of $csqf\,(\chi^2\,;\,\nu)$

Cdf = Cumulative distribution function

d = number of defectives so far (in a seq-sample plan)

d-of-b = degree of belief

D = parameter associated with binomial distribution, see Eq. 3-10

E_i = an event

h = parameter of accept and reject decision lines, see Eq. 3-12

H = new test result

H, L = subscripts: H implies Higher, L implies Lower

H_0 = null hypothesis

H_1 = alternate hypothesis

L, U = subscripts: L implies Lower, U implies Upper

m-sample = a multiple-sample plan

n = number of items tested so far (in a seq-sample) plan

N = sample size

N_i = sample size for interval i in multiplesample plans

OC = operating characteristic

pdf = probability density function

P_a = probability of acceptance

	= probability of event in $\{\cdot\}$
r	= number of failures
r^*	= number of failures for rejection
R	= s-reliability, Sf
$\bar{R}$	= s-unreliability, $\bar{R} = 1 - R$, a Cdf
RL	= reliability level
s, s_i	= slope associated with accept and reject decision lines; see **Eqs.** 3-12 and 3-28; number of test stations
s-	implies the word "statistical(ly)", or implies that the statistical definition is intended rather than the ordinary dictionary definition
seq-sample	= a sequential-sample plan
Sf	= Survivor function, $Sf\{\cdot\} = 1 - Cdf\{\cdot\}$ for a continuous variable
T	= total test time
T^*	= value of T for acceptance
T_M	= mission time, see par. 3-6.3
URL	= unacceptable reliability level
Wt	= waiting time before decision, see par. 3-6.3
z	= a standard s-normal variate, par. 3-8
α	= producer risk
β	= consumer risk
γ	= discrimination ratio
θ	= $1/\lambda$, exponential scale parameter, mean of failure time
λ	= constant failure rate, Poisson rate parameter
μ	= mean
$\chi^2_{P,\nu}$	= value of chi-square such that $csqf(\chi^2_{P,\nu}; \nu) = P$
1, 0	= subscripts: 0 implies H_0, 1 implies H_1
1-sample	= a single-sample plan

3-1 INTRODUCTION

After an equipment or system design is well established, reliability-testing changes from a design tool to a tool for making decisions and for determining if reliability goals have been met. These tests range from quality assurance tests, which are performed at the parts level on submitted lots, to reliability demonstration tests at the system level (Ref. 1).

A test program that produces spurious results can lead to accepting unreliable equipment. The cost of spare parts and extra maintenance can far exceed the cost of accurate testing. Therefore, a carefully designed test program which considers s-confidence levels, sample sizes, consumer and producer risks, and test cost must be developed. In any reliability test, many engineering factors must be considered in addition to statistical factors.

Often, it is difficult to perform a reliability demonstration test on a complete system, especially if the system is very large. For example, it would be almost impossible to perform a reliability demonstration test on a complete ballistic missile system, including missile, launcher, command and control, etc. In this case, techniques which permit system reliability (and the uncertainty therein) to be estimated from subsystem reliability data can be used.

3-2 CONCEPTS

This paragraph treats the statistical concepts for tests used to accept or reject an equipment or system, based on its degree of compliance with specified reliability goals.

3-2.1 TERMINOLOGY

The statistical basis of a decision test (Accept/Reject) is the theory of testing hypotheses. In reliability testing, the null hypothesis under test is: "the submitted lot or system conforms to the reliability requirement". *An* alternate hypothesis is also specified (or at least implied): "the submitted lot or system does not conform to the reliability requirement in some way". Rejecting the null hypothesis is often considered statistically equivalent to accepting the altemate hypothesis (Refs. 1 and 5), but other engineering considerations often apply.

When components are qualification tested, samples–rather than whole lots–are used. Therefore, the possibility of incorrect inferences due to sampling fluctuations has to be considered.

The following notation is used:

H_0 ≡ the null hypothesis that the product conforms to the reliability requirements (e.g., the mean life is equal to a specified value)

H_1 ≡ the altemate hypothesis that the reliability of the lot is at some reliability level considered to be unacceptable.

Incorrect inferences are of two types:

1. Type I Error: H_0 may be rejected when it is true. (Producer risk)

$\alpha \equiv$ Probability of Type I error, producer risk.

2. Type II Error: H_0 may be accepted when H_1 is true. (Consumer risk)

$\beta \equiv$ Probability of Type II error, consumer risk.

The probabilities of making these errors depend on the sample size and decision criteria. Table 3-1 summarizes the relationships.

It is easy to get boxed in by statistical terminology and assertions. For example, a valid conclusion of some statistical tests (not mentioned so far) is that there were not enough data to detect a difference–if it exists. *An* engineer can always call for more tests. It may mess up some previously calculated s-confidence or s-significance levels, and it may make new ones virtually **impossible** to calculate; but those levels are not the be-all and end-all of testing. The prime purpose of testing is to be sure only good equip-

TABLE 3-1

RELATIONSHIPS BETWEEN TEST DECISION AND TRUE SITUATION

Test Decision	True Situation: H_0 True	True Situation: H_1 True
Accept H_0	Correct Decision Probability = $1-\alpha$	Type II Error Probability = β
Accept H_1	Type I Error Probability = α	Correct Decision Probability = $1-\beta$

ment gets out in the field; don't lose sight of that goal—even at the expense of annoying a statistician.

In acceptance sampling, the probability of a Type I error α is commonly called the producer risk, since it represents the **risk** that a product conforming to the specification will be rejected. The probability of a Type II error β is called the consumer risk, since it represents the risk of accepting **a** product that ought to be rejected [$(1 - \beta)$ is known as the power of the test].

Reliability can be measured by various parameters, such as probability of survival, mean life, and failure rate.

The following nomenclature is employed:

1. *Reliability Level (RL)* is the level of the reliability-measure which the lot or system actually has.

2. *Acceptable Reliability Level* ***(ARL)*** is the ***RL*** considered to be acceptable, and represents the null hypothesis,

$$H_0: \quad RL = ARL \; . \tag{3-1}$$

(The Acceptable Quality Level, *AQL,* is the analogous term for acceptance tests based on fraction defective.)

3. *Unacceptable Reliability Level* ***(URL)*** is the *RL* considered to be unacceptable, and represents the alternate hypothesis,

$$H_1: \quad RL = URL \tag{3-2}$$

(The Lot-Tolerance-Percent-Defective **(LTPD)** is the analogous term for acceptance tests based on fraction defective.)

4. *Discrimination Ratio* γ *is a* ratio involving *ARL* and *URL,* and always is defined *so* that it is greater than one.

a. For mean-life requirements,

$$\gamma \equiv \theta_0/\theta_1 \tag{3-3}$$

b. For failure-rate requirements,

$$\gamma \equiv \lambda_1/\lambda_0 \tag{3-4}$$

c. For survival-probability requirements,

$$\gamma \equiv \bar{R}_1/\bar{R}_0 \tag{3-5}$$

where

θ = mean life (exponential parameter)

$\lambda = 1/\theta$ = constant failure rate

$\bar{R}$ = s-unreliability

the "0" subscript signifies the *ARL*

the "1" subscript signifies the *URL*

The *ARL* has a **high** probability of acceptance and the ***URL*** a low probability of acceptance. It is important that specified values of the *ARL* and *URL* be consistent with operational requirements. Generally, the ***URL*** is that value which is minimum for satisfying operational requirements, while the *ARL* is a highly acceptable value; the latest revision of Ref. **25** ought to be consulted for contractual situations.

This method of describing a sampling plan can be very misleading to a nonstatistician. It **is** probably best to consider this procedure **as** a means of picking two points on the Operating Characteristic (see par. **3-2.4)** of the sampling plan, in order to generate the complete Operating Characteristic. Once the entire Operating Characteristic is generated, one can forget the names of the two points used to generate **it**.

Many pairs of points will lead to the same

Operating Characteristic, even though the points look quite different. Once the entire Operating Characteristic has been generated, any pair of points which lie on the Operating Characteristic will give the same plan, even though the *ARL* and *URL* are completely different.

3-2.2 CONSUMER AND PRODUCER RISKS

The α and β risks represent the specified decision errors associated with nominally good and bad product, respectively. Since lower risks require more testing, a balance between the amount of test effort and the cost of a wrong decision is required (Ref. 1).

In general, for tests in the development stage, the α and β risks may be high, say 20%, and ought to be low for subsequent tests, such as production-acceptance tests, say 5% to 10%. This is so because the early tests usually are scheduled to allow later design changes, and the user at this point is concerned primarily with being assured that the equipment reliability is not completely unacceptable.

The next tests are those scheduled immediately after the first production run. The α and β risks for these tests are usually lower than for the development tests, since now the equipment design is complete and more equipments are available for test. Risks of 10% often are specified.

The third phase of testing is a sampling during production to ensure that acceptable quality is maintained. If the producer is continuously meeting the reliability goals, a limited amount of testing is desired, and the α and β **risks** can be higher than for the initial production tests.

3-2.3 *s*-CONFIDENCE LEVELS

Test requirements can be specified in terms of s-confidence levels. Such a specification, however, must be framed carefully. The specification ought to depend on the operational requirements of the system in which the components are located. For example, consider the following specifications for mean life θ:

1. Compute the 90% lower s-confidence limit θ_L. Since one can be 90% s-confident that the true mean life is greater than θ_L, if $\theta_L \geqslant 100$ hr, accept the lot; otherwise, reject it.

2. Compute the 90% upper s-confidence limit θ_U. Since one can be 90% s-confident that the true mean life is less than θ_U, if $\theta_U \leqslant 100$ hr, reject the lot; otherwise, accept it.

Specification No. 1 is equivalent to one in which the consumer **risk** is 10% at a true mean-life of 100 hr. Specification No. **2** is equivalent to one in which the producer risk is 10% at a true mean-life of 100 hr. The difference between the two tests is apparent. With Specification No. 1, only 10% of lots with a mean life of 100 hr will be accepted; while with Specification No. **2,** 90% of the lots with this same mean-life are accepted.

For most tests, the magnitude of α and β and the number of test observations N are related, so that specifying any two of the quantities determines the third. One approach, for nonsequential tests, is to specify α and N and to choose a test which minimizes β. For acceptance testing, the trend now is to specify β instead of α. If it is important that both α and β be specified, the sample size is completely determined. Because N is discrete, the exactly chosen values of α and β are not usually attainable. The α and β are then moved around, for suitable values of N, until an acceptable set is found. In sequential sampling, α and β are specified in advance, and the sample size is a random variable whose value is not predetermined but changes in successive tests; see pars. 3-5, 3-6, and 3-7.

Conventional wisdom gives more importance to the 2 special points *(a,ARL)* and *(β, URL)* than is justified. What is important is the entire function which relates true quality to accept probability over the entire range of possible quality. Par. 3-2.4 explains this function and its uses.

3-2.4 OPERATING CHARACTERISTIC CURVE

When two of the three quantities—N, α, and β—are specified, the accept-reject criterion of the acceptance test uniquely is determined for a given family of tests. It is then possible to generate the Operating Characteristic (OC) curve of the test plan. This curve shows the probability of lot acceptance over all possible incoming reliability levels. Two points on the *OC* curve are already determined—the a and β points with their corresponding reliability levels, *ARL* and *URL*, respectively.

For example, if the specification is in terms of a survival probability for a given period of time, the general shape of the OC curve is as shown in Fig. 3-1.

The probability of acceptance is a binomial probability parameter. It can be interpreted as the long-run proportion of lots that will be accepted. If, for example, the OC curve shows that a lot with a reliability of 0.80 will be accepted with a probability of *65%*, then in the long run *65%* of all lots submitted with 20% defective items will be accepted.

Each sampling plan has its own OC curve. This entire curve is affected by a change in sampling specifications.

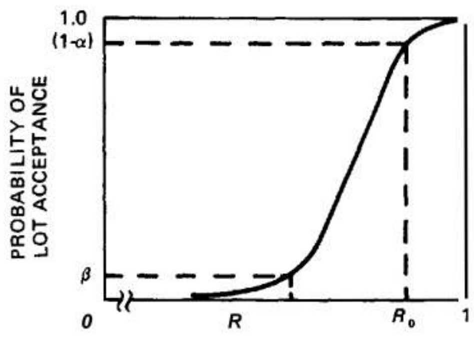


Figure 3-7. Typical Operating Characteristic Curve for Reliability Acceptance Test, H_0: $R=R_0$, H_1: $R=R_1$, ***Specified α and β (Ref 1)***

3-3 PRELIMINARIES TO TESTING

Test designers must consider many factors in addition to statistical ones. These engineering factors include test environment, equipment, and test procedures which must be defined in advance of the test and carefully controlled in order to ensure valid test results (Ref. *2)*. Several areas must be considered for a reliability test to succeed in providing adequate data:

1. Use of existing information
2. Selection of test parameters
3. Test procedures and instructions.

3-3.1 USING EXISTING INFORMATION

When designing a test, it is often impossible to include all possible levels of important factors influencing reliability. For example, if the three environmental factors—temperature, vibration, and humidity must be considered at four levels each—81 different combinations would have to be generated for a full-factorial design. If variations in input power, frequency, and similar factors had to be generated, the total number of combinations would be astronomically high. If historical information is available on tests made on similar equipments, it may be possible to eliminate some environmental combinations from the test. A simple example of this follows. If the interaction of temperature and vibration has been shown to be a critical factor only for high temperatures, then combinations of temperature and vibration for low temperatures can be ignored.

Since new equipment designs usually make use of many standard parts, circuits, and assemblies, it may be possible to use published information on their performance and reliability characteristics to reduce the amount of reliability testing required. These data must be evaluated carefully. A checklist of factors to consider in evaluating such data is given in Table 3-2. This list is to be used together with the engineer's knowledge of the system to determine how the existing data can be used. But remember, engineers tend to be overly optimistic about their designs.

The number of environmental combinations which must be applied to a system during reliability or to its components during qualification testing depends on how much information can be derived from reliability tests performed during research and development. These tests provide data that substantially reduce the number of combinations of environmental and other factors which must be considered in later testing.

3-3.2 SELECTION OF TEST PARAMETERS

The reliability parameters which are measured by a test are usually those which have been specified in the original contract and specifications (Refs. 1 and 2). Parameters such as Mean Life, MTBF, Reliability with Repair, and Availability, are specified as system measures of effectiveness. Usually, the system specification states that the system must meet some minimum level of a parameter at a specified s-confidence level. For example, the specification might state that the system must meet a 1200 hr minimum MTBF with a 90% s-confidence level. Various measures can be selected, depending on the system design, performance characteristics, and mission.

There are no formal rules for selecting the measures to be tested. Engineers and contracting officers must use their best judgment based on their experience when specifying test parameters. Considerable care must be devoted to the selection of parameters to be measured on the test because the types of test rigs, test equipment, and procedures to be used depend on the parameters tested. If the proposed test is too severe and expensive,

TABLE 3-2

BASIC CHECKLIST FOR DATA REVIEW

1. What is the source of the data?
2. When and where were the data obtained?
3. What was the purpose of the experiment?
4. Concerning which population can conclusions be drawn?
5. What was the experimental design?
6. How were sample items selected?
7. What were the operating and environmental conditions?
8. If failure rates are given, how was failure defined?
9. If repair rates are given, under what conditions were repairs made?
10. Is there any measure of experimental error?
11. What was the basis for computing time?
12. What type of test instrumentation was employed?
13. What are the major differences between items tested and those under consideration?
14. How do the results compare with those of similar investigations?

it may be deleted by a contract modification when calendar time and funds begin to run low.

3-3.3 TEST PROCEDURES

A detailed set of test procedures must be developed for quality assurance tests and reli- -ability demonstration tests (Refs. 1 and 2). The procedures are a crucial part of the test and must be carefully designed. The test procedures ought to include the following:

1. Purpose of test

2. Test items—description and sample selection

3. Test monitoring and review procedures

4. Test equipment required

5. Test equipment calibration procedures

6. Test equipment proofing

7. Environmental conditions to be applied

8. Operating conditions

9. Test-point identification

10. Definition of limits of satisfactory performance

11. Procedures for conducting test

12. Test report procedures and documents.

3-4 EXPERIMENTAL DESIGN

Reliability testing usually is performed on a small sample of items. Based on the test results, inferences are made about the population from which the sample was drawn. Principles of experimental design have been developed so that valid inferences can be drawn. These principles are briefly discussed in this paragraph (Refs. 14).

3-4.1 THE POPULATION

In experimental design, the population is the set of objects about which inferences are to be drawn (Ref. 1). Usually the population is an actual rather than an abstract group. In equipment reliability testing, however, inferences must be made about items not actually in existence at the time of the test, but which will be produced at some time in the future. For example, when a demonstration test is to be performed on a complex equipment, only three or four of the equipments may actually exist and only one or two may be used for the test. It is assumed that if the equipments are manufactured under the same production processes at a future time, then reliability inferences based on tests of small numbers of equip ments are valid. If a major redesign or retrofit takes place as a result of field experience, the results of previous tests may not be valid and additional testing may be required.

There are no formal rules for dealing with the problem of population definition, except that it is necessary to define as completely as possible the equipment test procedures and test conditions from which inferences are to be made.

The method of drawing the random sample determines the population about which inferences are to be made. Ask the question: "From what population is my sample a truly random sample?" Generally, the population about which inferences legitimately can be made is much more restricted than the engineer would like. This consideration is extremely important in evaluating much of the theoretical/empirical research in reliability.

3-4.2 ELIMINATION OF BIAS

Bias in a reliability test may result in false conclusions. Bias may arise through the choice of sample, experimenter influences, instrument errors, physical or laboratory variations, or the experimental design itself. Bias can be reduced through experimental controls and randomization (Ref. 1).

3-4.2.1 Experimental Controls

In part qualification tests, parts used for control are selected from the same population as the test sample and are subjected to the same conditions with the exception of the conditions that are being studied (Ref. 1). Controls permit the test engineer to determine if the test results are a function of the conditions applied or of other variables that are biasing the results. The test engineer may be interested, for example, in the variation of resistance with time when resistors are subjected to a specific set of operating and environmental conditions. If control resistors are not used, the engineer cannot be sure if observed fluctuations in resistance are time fluctuations, or are caused by the load, or are due to a combination of time and load. A group of similar resistors that did not have any load imposed on them could be used as a control.

The use of controls is more difficult to implement on reliability demonstration tests because only a limited number of equipments or systems are tested. If only a few systems are available, it makes more sense to use them for demonstration testing rather than to waste a system as a control.

3-4.2.2 Randomization

The test engineer can use randomization procedures to protect against the introduction of bias. An example of a situation in which randomization is required is a life test experiment designed to test the effects of power dissipation and temperature on resistors (Ref. 1). A group of resistors is to be placed in an oven which maintains a constant temperature. The resistors are to be split into three groups, each with a different power dissipation. If there are three racks in the oven, the simplest procedure is to place each power dissipation group on a rack.

During the life test, hot spots may develop in the oven, especially since different levels of power dissipation generate different amounts of heat. Also, because hot air rises, the top rack could be somewhat hotter than the bottom rack. Placement of all resistors of one group on the same rack might lead to biased results. To avoid this possibility, the oven location of each resistor can be determined in a random manner, such as through the use of a random-number table. Then the unknown or uncontrollable effects of temperature variation within the oven would be distributed within all power dissipation groups (bias is converted into an uncertainty which good statistical analysis can estimate and correct for).

In reliability demonstration tests, the order and combination of applied stresses and environments can be determined randomly if they are not otherwise specified. This ensures that all such combinations are equally likely on the test and that no biases are introduced.

3-4.3 EXPERIMENTAL UNCERTAINTY

Experimental uncertainty is the random effect of factors over which the experimenter does not have complete control (Ref. 1). Two important sources of experimental uncertainty are (1) the inherent variability in manufactured parts, and (2) the random variation in the physical conduct of experiments.

Experimental uncertainty is associated with the concept of precision, the repeatability of results. The larger the experimental uncertainty, the less precise the results. The size of the experimental uncertainty can be decreased through replication of the experiment. Replication is the term used to indicate the number of parts with which a specific test condition is associated. The total sample size is the sum of all the replications. For example, if a test is conducted at A levels of temperature, B levels of vibration, and C levels of acceleration, there are a total of A•B•C combinations. If each of these combinations has R replications, the total number of parts to be sampled is (A•B•C)R. Increasing the number of replications will increase the precision of the test, i.e., reduce the experimental uncertainty.

A valid test must provide a good estimate of the experimental uncertainty. Randomization provides the test engineer with a technique that ensures good estimates of uncertainties. To support the results of an experiment with probability statements, randomization is necessary in the sample selection and in the experimental design.

Standard deviations and s-confidence statements are two popular, good ways of measuring uncertainty.

3-4.4 SAMPLE SELECTION

The elimination of possible sources of bias is an important consideration in the choice of sample items. Before a sample selection scheme is chosen, the population from which the sample is to be drawn must be precisely and correctly defined. If resistors are the parts under consideration, the test engineer must ask if the population is to contain all resistors, or just carbon composition; all values of resistance; S%, 10%, or 20% tolerance limits; and so on. The population must be precisely and correctly defined, since conclusions of the experiment are limited to that population from which the actual sample is truly a random one. In demonstration tests, "identical" systems must be tested (Ref. 1).

The principal requirement of any sampling procedure is that it yield representative samples. A representative sample is a miniature of the population. To make inferences about the population from the sample results, the sample selection must be random. The simplest kind of random sample results when every item in the population has an equal chance of being chosen for the sample. A simple random sample does not necessarily result in a representative sample. Stratified sampling can be used to obtain a sample that is representative of the universe of items to be tested. However, within each stratum, the final selection procedure is random.

In system demonstration testing, the Sample selection is limited to a few systems. Often, a random sampling procedure cannot be followed because the testing is limited to the first few preproduction models. Then it is important to distinguish between the special attention preproduction models are given (from purchased parts to final inspection) vs the kind of actual production system in the future. For example, how will the quality of incoming steels be checked to be sure they meet the specifications, month after month? What will long term variations in all kinds of properties do to estimates of standard deviations which were based on a few samples?

3-5 TYPES OF TESTS

Tests often are classified as "attributes" or "variables".

An attributes test is one in which each item is tested, and judged to be a success or a failure. Attributes tests usually are used for testing a 1-shot item in which time or cycles are not involved. It is possible, however, to include time by testing each item for a specified period and counting the number of successful and failed items. This type of test is not usually called a life test, because the time at which the failures occurred during the testing period is not considered.

A variables test is one in which some characteristic of the test items is measured on a continuous scale, such as amplitude, power output, or life. If the characteristic is not life, and if each item is tested for a specified time, the characteristic is measured at the end of this time.

A life test is a variables test in which the important, measured thing is the time it takes an item to reach a particular condition, usually a "failure". The times at which the failures occur are recorded.

The kind of test which should be used often can be decided by evaluating the application of the item and its reliability objective. If the reliability objective is stated in terms of probability of surviving a fixed mission, an attributes test at the end of the mission often is used. If the goal is in terms of time to failure, a life test usually is indicated.

When more than one type of test could be employed; such factors as type of information provided, degree of protection afforded, amount and cost of inspection, and ease of administration should be considered. Table 3-3 summarizes the conventional wisdom on advantages and disadvantages of each type of plan with respect to these considerations.

3-5.1 SINGLE- OR MULTIPLE-SAMPLE PLANS

The terminology is not standard, and so it will be explained here. All samples are presumed to be random from a large lot.

Single-sample (1-sample) plans test one sample of items, and then decide to accept or reject the lot. The analysis usually is exact.

Multiple-sample (m-sample) plans can test several samples of items. After every sample, except the last, the plan decides to accept or reject the lot, or to take another sample. After testing the last sample, the plan decides to accept or reject the lot. The analysis usually is exact.

Sequential-sample (seq-sample) plans are, in effect, m-sample plans whose boundaries have been drawn differently. Most of the analyses are very approximate and deal with untruncated plans (sample size $\to \infty$). Naturally, all real plans are somehow truncated, Refs. 21 and 22 show exact analyses of several plans and give references to exact analyses of other plans.

Fig. 3-2 is a pictorial description of sampling plans. The illustrations are all for attributes plans, but the axes can be relabeled for variables plans. The course of the test can be represented as a path, a series of points, wherein the cumulative result of testing each item is plotted. When the path touches or crosses a boundary, the decision appropriate to that boundary is made. Testing continues until a boundary is reached.

Fig. 3-2(A) is the traditional 1-sample plan. The boundary is a rectangle. The slash shows the dividing line between accept and reject; the point (N,c) is on the accept boundary; the point $(N, c + 1)$ is on the reject boundary.

Fig. 3-2(B) is the traditional 2-sample plan. The point $(N_1 + N_2, c_2)$ is "accept"; the point $(N_1 + N_2, c_2 + 1)$ is "reject". A traditional r-sample plan would have r such rectangles put one on top of the other. The accept line is a series of steps; the reject line is horizontal and even with the top of the last step. Very seldom is r greater than 2 in a widely used plan, and almost never is it greater than 3.

Fig. 3-2(C) is the traditional sequential-sample plan. The shape of the truncation is

TABLE 3-3

COMPARISONS BETWEEN ATTRIBUTES AND VARIABLES TESTING

Factor	Attributes	Variables (other than life)	Variables (Life Test)
Use of Item	Single operation.	Single operation.	Repetitive or continuous operation over time.
Type of Information Yielded	Number or fraction of sample that failed to meet specified quality characteristics at a given point in time.	Distribution of some quantitative output at a given point in time. Provides most information for quality improvement.	Distribution of failures over time.
Reliability Goal	Fraction-defective or probability-of-survival over a fixed time period.	Output tolerance limits which define success or failure possibly applying after a fixed period of operation.	Mean life, failure rate, or probability of survival for a fixed time period.
Sample Size for Given Protection	Usually highest.	Usually lower than attributes test for corresponding plan.	Lower than attributes test for corresponding plan.
Ease of Inspection	Requires relatively simple test equipment and less-qualified personnel.	More complex test equipment and better trained people required than for attributes tests.	Continuous observation necessary for most types of tests. Highly trained people required. Difficult to maintain controlled test conditions.
Simplicity of Application	Data recording and analysis is fairly simple. Single set of attributes criteria applies to all quality characteristics.	More clerical costs than attribute plans. Variables criteria needed for each quality characteristic.	More clerical costs than attribute plans. Has one set of criteria for all quality characteristics.
Statistical Considerations	No assumptions on failure distribution required. Binomial distribution applies for most cases. Extensive tables are available. (Refs. 5-7)	Often requires a parametric assumption on the distribution of the characteristic considered. (Ref. 8)	Usually requires an assumption of a time-to-failure distribution. Tables available for exponential and Weibull distributions. (Refs. 9 and 10)

See Ref. 19 for nonparametric tests and Ref. 20 for a general discussion of reliability and life tests.

d = defectives found so far

n = number tested so far

c = acceptance number

A = "Accept" region

R = "Reject" region

The dotted line is $d = N + 1$, and is unreachable.

All boundaries, for discrete variables, are a series of points.

The d and n scales are not the same.

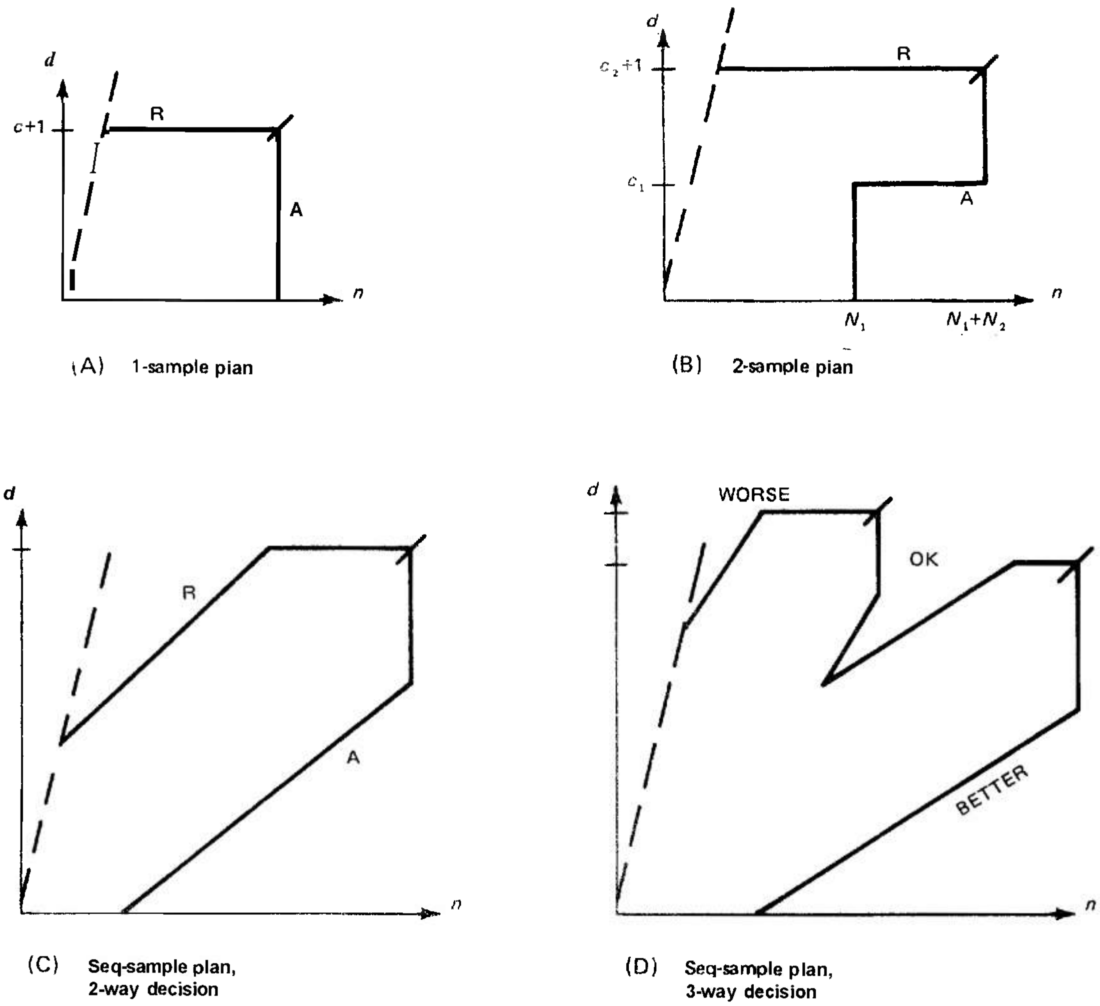


*Figure 3-2. Various Sampling **Plans***

usually optional, and either the horizontal or vertical segment can be missing.

Fig. 3-2(D) is an unusual sequential-sample plan; it has 3 decision regions on the boundary. For example, in the "worse" region, one might reward the supplier. In the supplier in some other way. In the "better" region, one might reward the supplier. In the "OK" region, the lot would be accepted but with no penalty or reward.

For each plan, the OC curve can be derived (in principle anyway, in practice it might be difficult, expensive, and/or tedious). It will show the probability of a particular decision vs true quality. *Also,* the Average Sample Number **(ASN)** can, in principle, be derived for each plan, as a function of true quality.

It is easy enough to estimate the actual quality of the lot, but it is quite difficult to find the probability distribution of that estimate unless the sample size is fixed. In all the plans shown in Fig. 3-2, the sample size is a random variable because a decision can (must) be made any time a boundary is touched (this is called curtailed-sampling). Since this chapter deals only with decision making, not estimation, the difficulties in estimating will not be considered further.

There is no rule or law of statistics (or anything else) that determines what shape the decision boundary must have. Some shapes may be better than others, according to specific criteria, but no shape (or implied plan) is always better in all ways than any other plan.

Two examples are given for the usual m-sample plans. Notation follows:

c_i = acceptance number, i.e., if $d \leqslant c_i$ when $n = \sum_{j=1}^{i} N_j$ then accept the lot.

If $d = c_m + 1$, reject the lot. Otherwise continue sampling.

N_i = size of sample i

m = maximum number of samples

n = number tested so far

d = number of defectives found so far

1-sample Plan

$N_1 = 100, c_1 = 3$

This means that if $n = 100$ and $d \leqslant 3$, accept the lot, if $d = 4$, reject the lot. Often the simplest administrative instructions are presumed to be: Test the whole lot of 100; if "$d \leqslant 3$", accept the lot, otherwise reject it.

2-sample Plan

$N_1 = 100, c_1 = 3$

$N_2 = 200, c_2 = 7$

This means that if $n = 100$ and $d \leqslant 3$, accept the lot. If $n = 100$ and $3 < d \leqslant 7$, take a sample of 200 more. If $n = 100 + 200 = 300$, and $d \leqslant 7$, accept the lot. If at any time, $d > 7$, reject the lot. As in the 1-sample plan, it is often thought to be simpler not to have the inspector concern himself about the number of defectives, except at $n = N_1$ and $n = N_1 + N_2$.

In the seq-sample plan, it is usually considered that a decision of some sort is made after testing each specimen. In practice this need not be true. When d is discrete, it is possible at the beginning to determine the minimum n ($n = n_1$) for acceptance (i.e., $d = 0$) and to take a sample of that size. If the point (n_1, d) is within bounds, the minimum "additional sample size to accept" can be calculated, and a sample of that size drawn. This process is repeated until a decision boundary is reached.

Conventional wisdom attributes smallest sample sizes to seq-sample tests, largest sample sizes to 1-sample tests, and in between to m-sample tests. However it is difficult to compare sample sizes because:

1. Sample size can be a random variable; usually then one deals with **ASN.**

2. The OC curves for the tests that are being compared are not (and cannot be) exactly the same (see pp. 261-262 of Ref. 23), even though they may be close in some regions of true quality.

3. The ASN is a function of true quality. It has been shown (see p. 262 of Ref. 23) that the usual seq-sample test of Wald, has an **ASN** no larger than any other sample plan at the 2 true-quality points where the plans intersect (often these 2 points will be the consumer and producer risk points).

If the distribution of true-quality of incoming lots is known rather well, then consult a statistician to optimize the sampling procedure (of course, choosing the criteria for optimization will be an exciting task in itself).

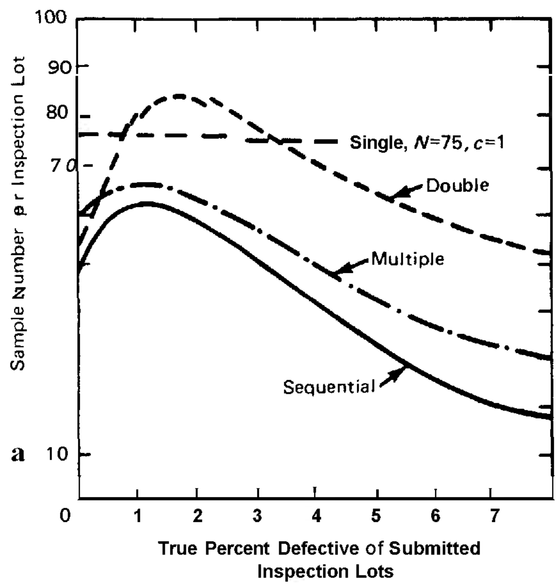
Fig. 3-3 compares the ASN for several sampling plans. The usual procedure is used for the 1-sample and 2-sample plans (i.e., d is monitored only at the possible acceptance points). In Fig. 3-3 the plans are asserted to be roughly equivalent to the 1-sample plan of $N = 75$, $c = 1$.

Table 3-4 compares some characteristics of sampling plans.

3-5.2 TRUNCATION

A truncated life test is one in which testing is terminated after a random variable reaches a preassigned number, e.g., number of failures or number of test hours. For practical reasons, all life tests are truncated because of economic and scheduling factors. Truncated life tests are especially suitable when the failure rate is constant. If the failure-time distribution is s-normal or lognormal, however, the mathematical difficulties of evaluating the results of truncated tests are quite formidable. This is also true for other failure distributions with nonconstant failure rates that involve more than one unknown parameter.

Seq-sample tests ought not to be used in contractual situations unless the OC curve



Notes: The 2 points on the OC curve, which are nominally the same for all plans, are:

(1) Producer risk 10%, at 0.7% defective

(2) Consumer risk 10%, at 5% defective.

See Ref. 1 for details of analysis.

Figure 3-3. Average Amount of Inspection Under Single, Double, Multiple, and Sequential Sampling (ASN Curves)[1]

TABLE 3-4

COMPARISON OF SINGLE, MULTIPLE, AND SEQUENTIAL SAMPLE PLANS
(Adapted and modified from Ref. 1)

Characteristic	1-sample	m-sample	seq-sample
Sample Size	Known (can be a random variable)	Average can be computed for various incoming quality levels. Often less than 1-sample.	Average can be computed for various incoming quality levels. Often less than m-sample.
Decision Choices	Accept or reject	Accept, reject, or take another sample until final sample is selected	Accept, reject, or test another item
Predetermines Characteristics	Two of the three quantities N, α, or β	Same as 1-sample	Fix α and β; N is a random variable
Statistical Considerations	Must know distribution of sample statistic	Same as 1-sample	Same as 1-sample, α, β are rarely known exactly. Usual formulas are very approximate.
Personnel Training	Requires least training	Better trained people required than for single	Requires most training
Ease of Administration	Easiest. Scheduling can be fairly precise and precise test-cost estimates can be made	More difficult than 1-sample since the exact number of rests is unknown. Only average test. costs can be estimated	Most difficult in terms of testing, scheduling, and overall administration. Most time consuming.
Miscellaneous	Best used for testing situations where ease of administration is most important and cost of testing is relatively unimportant.	Has psychological advantage in that supplier is given a "second chance" by taking further samples if first sample results indicate a marginal lot.	Can be most "efficient" under many circumstances. Can be truncated and still maintain good α, β. Consult statistician knowledgeable in this area.

has been determined for the particular tests. The nominal OC curve can be off by a factor of 2 or so in α and β. See Refs. 21 and 22 for techniques of analyzing the seq-sample plans.

3-5.3 SPECIAL TESTS

When the "reject" decision is equivalent to "test the entire lot" rather than "destroy the lot" or something equally as drastic, the test described in the following paragraph may have advantages for the producer, while retaining protection for the consumer.

Use an m-sample or seq-sample test, but move the "reject" part of the boundary to "infinity". This means that a lot will never by rejected; in lieu of that, one may have to test the entire lot before accepting it. But that is what a "reject" decision would have meant, anyway. Details of such plans are not readily available in the open literature. Ref. 26, sections B.0933 and B.0935 discuss such plans, at least in principle.

An alternative is to put the "reject" line very high; so the probability of fully testing an acceptable lot is very small.

3-5.4 ASSUMING A FAILURE LAW

In life tests, a failure distribution is almost always assumed, and it usually is the constant failure-rate distribution. The parameter to be judged is the failure rate (or, equivalently, its reciprocal).

If one just considers the fraction good or bad, then the parameter to be judged is the fraction bad (or, equivalently, the fraction good). This assumption (the parameter "fraction bad" is the same for every item tested) is much more likely to be fulfilled than is the assumption of a specific failure distribution.

Conventional wisdom states that parametric tests are "better" than nonparametric tests since, for a given amount of testing, more precise estimates are obtained from the parametric tests. However, see Ref. 19 for a more complete discussion of this point; many nonparametric tests are very "good", and do not have the big disadvantage of having made the wrong distributional assumption.

Many specifications are written in terms of parametric testing or in terms of such properties as mean-life which are not too suitable for nonparametric tests. For example, nonparametric tests of central tendency apply to the *Cdf* while the specification may be in terms of mean life. This might require a change in specifications.

An incorrect assumption of the underlying failure distribution in a parametric test can lead to an OC curve that differs greatly from that planned, especially for small sample sizes. This is not a problem in nonparametric tests. Also, nonparametric tests are generally easy to conduct and evaluate, and often require only counting, adding, subtracting, or ranking; see Ref. 19.

3-5.5 REPLACEMENT

Replacement here is not used in its statistical sense of discrete nondestructive sampling (e.g., urn problems or decks of cards).

Replacement tests are those in which failed items are replaced by new items, so that the test stations are always full. If the items are complex, replacement may be interpreted to be restoration of the failed item to new condition by repair or replacement of failed components. In nonreplacement tests, failed items are not replaced or repaired; therefore, the number of items on test decreases as life testing progresses.

Generally, the effect of using a replacement test is to decrease the calendar waiting time before a decision can be made compared to that of a nonreplacement test with the same number of items on test originally. This savings in calendar time is accomplished at the cost of having to place more items on test. If a sequential test is used, it is usually preferable to plan for a replacement test, since all items may fail in a nonreplacement test before a decision is made, and more test items will have to be obtained.

This is a practical engineering problem, not a statistical one. The required data are the actual operating life of each item, regardless of when it was put on test.

3-5.6 ACCELERATED LIFE TESTS

An accelerated test is one in which the test conditions are adjusted to accelerate failure, i.e., to be more severe. While accelerated tests can be used to discover and evaluate critical weaknesses in the parts or design, their attractiveness in acceptance tests is that the amount of test time is reduced since the required number of failures for a decision will occur relatively early. This reduction in waiting time is most important for items that have very high reliability goals, because the amount of test time required to establish conformance can be very large.

If the stress conditions are accelerated, the reliability goal under standard stress conditions had to be modified accordingly. Therefore, the relationship (approximate) of reliability to the acceleration factor must be known, so that appropriate test criteria can be established. Most accelerated life tests are performed at the part level because of the stringent reliability requirements existing at this level and because stress/failure relationships are relatively easy to determine through experimentation.

It pays to repeat the warnings about extrapolation from experimental data. If a regression line has been determined for severity-level vs parameter of a distribution (usually the constant failure rate), the uncertainty in extrapolated value ought to be determined at the nominal conditions. The actual failure rate can easily be uncertain by a factor of 10! See Ref. 24 for the case where the time-independent failure rate obeys the Arrhenius temperature equation.

It is easy to be careless about just what is being accelerated in an accelerated test. Most often, it ought to be a particular parameter in a life distribution. If the distribution has more than one parameter, then the acceleration behavior of all parameters ought to be justified. For this reason, accelerated tests are more useful in the design/development stage. There, it is not the actual life that is important, but the failure mode/mechanism itself—and whether it is likely to occur at usual conditions.

An important consideration in using an accelerated test for acceptance testing is that the supplier might design his product to pass your accelerated test, rather than to do well in the field.

Further details on accelerated tests are given in Chapter 7 "Accelerated Tests".

3-6 BINOMIAL PARAMETER

This paragraph describes the statistical characteristics of tests to which binomial distribution theory applies. For reliability, the test characteristic of interest is often the fraction-defective (failed) over some fixed time of operation.

If $\bar{R}_0$ represents the *ARL* and $\bar{R}_1$ the *URL*, the test specification is of the following form:

$$H_0: \quad \bar{R} = \bar{R}_0 \qquad (3\text{-}6a)$$

$$H_1: \quad \bar{R} = \bar{R}_1 \qquad (3\text{-}6b)$$

and

Producer **Risk:** α
Consumer **Risk:** β

For "1-shot" items, where time is not involved, the item is tested for performance without considering test time *per se.*

3-6.1 1-SAMPLE

If the lot size is large in relation to sample size, the binomial distribution can be used to generate the OC curve of a 1-sample plan. Accept or reject decisions are made by testing N items for time T.

The lot is accepted if the number of failures is less than or equal to c, the acceptance number. If R is the true s-reliability for the test mission, the probability of acceptance P_a (a point on the OC curve) is

$$P_a(\bar{R}) = \sum_{k=0}^{c} \binom{N}{k} \bar{R}^k R^{N-k} \tag{3-7}$$

where

N = the number of items tested

c = the maximum allowable number of failures (accept with c, reject with $c + 1$)

R = the true s-reliability

$\bar{R}$ = $1 - R$

If $\bar{R}$ is small enough, say $\bar{R} \leqslant 0.1$ and N is large enough, say $N \geqslant 10$, the Poisson approximation can be used to obtain P_a from

$$P_a(\bar{R}) = \sum_{k=0}^{c} \frac{e^{-\mu}\mu^k}{k!} = poif(c; N\bar{R}) \tag{3-8}$$

where

$\mu \equiv N\bar{R}$

See par. 2-3.1 for a discussion of this distribution.

In order to meet the test requirements, values of N and c must be chosen so that

$$\left.\begin{aligned} P_a(\bar{R}_0) &= 1 - \alpha \\ P_a(\bar{R}_1) &= \beta \end{aligned}\right\} \tag{3-9}$$

Because c and N are integers, it usually is not possible to find a 1-sample plan that satisfies Eq. 3-9 exactly. Presumably this has been realized before contractual obligations were incurred. Since the R_0, R_1, α, and β are somewhat arbitrary anyway, a reasonable set of parameters is chosen which is close to the original plan.

Table 3-5 can be used to determine the 1-sample plan (N and c) for various values of the discrimination ratio γ (see Eq. 3-5).

$$\gamma \equiv \bar{R}_1/\bar{R}_0$$

and various sets of α and β. For a given set of α, β, and γ, Table 3-5 lists c and a parameter D. The sample size N is

$$N = [D/\bar{R}_0] \tag{3-10}$$

where $[x]$ is the largest integer $\leqslant x$. Table 3-5 is an extension of Table 2C-5 in Ref. 8.

Example Nos. 22 and 23 illustrate the procedure.

Table 3-6 presents 1-sample plans that approximate the OC requirements for common sets of R_0, R_1, α, and β. The s-expected sample sizes for "equivalent" sequential plans are also shown (Ref. 1).

Tables 3-5 and 3-6 have some differences (usually minor) in N and c for the same nominal 1-sample test specification. Table 3-5 is based on the Poisson distribution, and the plans are derived so that α is guaranteed and β is no more than specified. Table 3-6 is based on the binomial distribution, and the criterion used was to meet both the α and β requirements as nearly as possible. If

[text continues on page 3-24]

TABLE 3-5

ATTRIBUTE 1-SAMPLE PLANS FOR NOMINAL a, β, γ (Ref. 1)

	$a = .01$						$\alpha = .05$						$\alpha = .10$					
	$\beta = .01$		$\beta = .05$		$\beta = .10$		$\beta = .01$		$\beta = .05$		$\beta = .10$		$\beta = .01$		$\beta = .05$		$\beta = .10$	
γ	c	D	c	D	c	D	c	D	c	D	c	D	c	D	c	D	c	D
1.5	135	110.4	100	79.1	82	63.3	94	79.6	66	54.1	54	43.4	76	66.0	51	43.0	40	33.0
2	45	31.7	34	22.7	29	18.7	32	24.2	22	15.7	18	12.4	25	19.7	17	12.8	14	10.3
2.5	26	16.4	20	11.8	17	9.62	18	12.4	13	8.46	10	6.17	14	10.3	10	7.02	8	5.43
3	18	10.3	14	7.48	12	6.10	12	7.69	9	5.43	7	3.98	10	7.02	7	4.66	5	3.15
3.5	14	7.48	11	5.43	9	4.13	9	5.43	7	3.98	6	3.29	7	4.66	5	3.15	4	2.43
4	11	5.43	9	4.13	8	3.51	8	4.70	6	3.29	5	2.61	6	3.90	4	2.43	3	1.75
4.5	10	4.77	8	3.51	7	2.91	6	3.29	5	2.61	4	1.97	5	3.15	3	1.75	2	1.10
5	8	3.51	7	2.91	6	2.33	6	3.29	4	1.97	3	1.37	4	2.43	3	1.75	2	1.10
7.5	5	1.78	4	1.28	4	1.28	3	1.37	3	1.37	2	.818	3	1.75	1	.532	1	.532
10	4	1.28	3	.823	3	.823	3	1.37	2	.818	2	.818	2	1.10	1	.532	1	.532

$P_a(\bar{R})$ = probability of acceptance, given the true unreliability is $\bar{R}$

c = acceptance number (lot is accepted for c or fewer defectives; it is rejected otherwise)

N = sample size

a, β = nominal producer and consumer risks (actual risks depend on the exact values of c and N)

D — D is defined such that $N = [D/\bar{R}_0]$, where $[x] \equiv$ "largest integer $\leqslant x$"

$/ = \bar{R}_1/\bar{R}_0$

$P_a(\bar{R}_0) \approx 1 - \alpha; P_a(\bar{R}_1) \leqslant \beta$

Example No. 22

A small electronic subassembly is subjected to a reliability test. The subsystem is tested for 200 hr and the number of failures is counted. The *ARL* is $\bar{R}_0 = 0.05$ with $\alpha = 10\%$. The *URL* is $\bar{R}_1 = 0.20$ with $\beta = 5\%$. Determine the sample size and acceptance number for a 1-sample attributes test.

Procedure	Example
1. Calculate the discrimination ratio from Eq. 3-5.	1. $\gamma = 0.20/0.05$ $= 4.00.$
2. Find c and D from Table 3-5.	2. For $\alpha = 0.10$, $\beta = 0.05$, and $\gamma = 4.00$, from Table 3-5, $D = 2.43$ $c = 4.$
3. Calculate the sample size, from Eq. 3-10.	3. $N = \left[\dfrac{2.43}{0.05}\right] = [48.6]$ $= 48.$
The test plan is $N = 48$, $c = 4$.	
4. Use Eqs. 3-7 and 3-9 to calculate the true α and β.	4. true $\alpha = 9.07\%$ true $\beta = 2.48\%$.

Example No. 23

Design a 1-sample plan so that if the average lot reliability for a 100-hr period is 0.99, there is a 90% probability of acceptance, and if the average lot reliability is 0.95, there is only a 10% probability of acceptance.

Procedure	Example
1. State the parameters of the problem.	1. The parameters are $\alpha = 0.10$ $\beta = 0.10$ $R_0 = 0.99$, $\bar{R}_0 = 0.01$ $R_1 = 0.95$, $\bar{R}_1 = 0.05$.
2. Compute γ from **Eq. 3-5.**	2. $\gamma = 0.05/0.01$ $= 5.00$.
3. Enter Table 3-6 for the appropriate parameters and determine N and c for a 1-sample plan.	3. For a 1-sample plan with $\alpha = 0.10$, $\beta = 0.10$, and $\gamma = 5$, from Table 3-6 we get $N = 110$ $c = 2$. Therefore, 110 items must be tested for 100 hr each, and the lot is accepted if 2 or fewer failures occur.
4. Use Eqs. 3-7 and 3-9 to calculate the true α, β.	4. true $\alpha = 9.87\%$ true $\beta = 8.3\%$.

TABLE 3-6

ATTRIBUTE SAMPLING PLANS FOR SOME NOMINAL α, β, γ (Ref. 1)

(The 1-sample and seq-sample plans will not have the same α, β; and neither plan will have the actual α, β.)

			1-sample plan		seq-sample plan Average Sample Number			1-sample plan		seq-sample plan Average Sample Number			1-sample plan		seq-sample plan Average Sample Number			1-sample plan		seq-sample plan Average Sample Number		
R_0	R_1	γ	N	c	R = I	$R = R_0$	$R = R_1$	N	c	R = I	$R = R_0$	$R = R_1$	N	c	R = I	$R = R_1$	$R = R_0$	N	c	R = I	$R = R_0$	$R = R_1$
			$\alpha = 0.10, \beta = 0.10$					$\alpha = 0.10, \beta = 0.20$					$\alpha = 0.20, \beta = 0.10$					$\alpha = 0.20, \beta = 0.20$				
0.99	0.98	2	950	13	199	437	582	650	9	136	285	451	650	9	188	339	379	400	6	125	206.9	275
	.97	3	320	5	104	175	142	180	3	71	114	110	220	4	98	136	93	140	3	65	83	67
	.95	5	110	2	53	69	43	60	1	36	45	33	78	2	50	54	28	60	2	33	33	20
	.90	10	37	1	23	25	12	30	1	16	16	9.4	22	1	22	19	7.9	16	1	15	12	5.8
.95	.90	2	190	13	40	103	87	113	8	28	67	67	129	9	38	80	56	78	6	25	49	41
	.85	3	60	5	20	35	25	35	3	14	23	19	46	4	19	27	16	31	3	12	16	12
	.75	5	20	2	9.3	12	7.8	11	1	6.4	7.9	6.1	16	2	8.8	9.4	5.1	11	2	5.9	5.8	3.7
	.50	10	8	1	3.4	3.6	2.2	5	1	2.3	2.3	1.7	4	1	3.2	2.8	1.4	4	1	2.2	1.7	1.0
.90	.80	2	80	11	19	48	40	56	8	13	31.3	30	59	8	18	37	26	39	6	12	23	19
	.70	3	25	4	8.8	15	11	18	3	6.0	9.9	8.9	23	4	8.3	11	7.5	9	2	5.5	7.2	5.4
	.50	5	9	3	3.7	4.8	3.4	5	1	2.6	3.1	2.7	8	2	3.5	3.7	2.2	5	2	2.4	2.3	1.6
.85	.70	2	49	10	11	29	25	33	7	7.7	18.6	19	35	7	11	22.1	16	21	5	7.1	14	11
	.55	3	16	4	5.1	8.6	6.8	11	3	3.5	5.6	5.4	10	3	4.8	6.6	4.5	6	2	3.2	4.1	3.3
.80	.60	2	33	9	7.6	19	17	24	7	5.2	12	13	24	7	7.2	15	11	16	5	4.8	9.1	8.0
	.40	3	9	3	3.2	5.3	4.6	6	2	2.2	3.4	3.6	7	3	3.0	4.1	3.0	4	2	2.0	2.5	2.2

$P_a(\bar{R}_0) \approx 1 - \alpha;\ P_a(\bar{R}_1) \leqslant \beta$

$P_a(\bar{R})$ = probability of acceptance, given the true unreliability is $\bar{R}$

c = acceptance number (lot is accepted for c or fewer defectives; it is rejected otherwise)

α, β = nominal producer and consumer risks, respectively (actual risks depend on the exact values of c and N)

N = sample size

the actual α and β were shown, the source and explanation of the difficulties would be obvious.

Tables 3-5 and 3-6 illustrate the following major points:

1. Sample size varies inversely with α, β, and γ ($\bar{R}_0$ fixed).

2. Sample size varies inversely with $\bar{R}_0$ (γ fixed).

3-6.2 SEQUENTIAL SAMPLING

In order to construct a seq-sample plan, the accept and reject decision lines (such as those shown in Fig. 3-2(C)) are computed from:

$$\left.\begin{aligned} a_A &\equiv \ell n \left(\frac{1-\alpha}{\beta}\right) \quad b_1 \equiv \ell n\, (\bar{R}_1/\bar{R}_0) \\ a_R &\equiv \ell n \left(\frac{1-\beta}{\alpha}\right) \quad b_2 \equiv \ell n\, (R_0/R_1) \end{aligned}\right\} \quad (3\text{-}11)$$

$$\left.\begin{aligned} s &\equiv \frac{b_2}{b_2+b_1} \quad \text{(slope)} \\ h_A &\equiv \frac{a_A}{b_2+b_1} \quad \text{(intercept)} \\ h_R &\equiv \frac{a_R}{b_2+b_1} \quad \text{(intercept)} \end{aligned}\right\} \quad (3\text{-}12)$$

Ordinarily, the a's, b's, h's, and s are all positive; s is always between $\bar{R}_0$ and $\bar{R}_1$. The accept line r_A equation is

$$r_A = -h_A + sn \qquad (3\text{-}13a)$$

The reject line r_R equation is

$$r_R = h_R + sn \qquad (3\text{-}13b)$$

where

n = the actual number of items tested so far

r = number of failures observed so far

The 3 decisions at each point are:

1. Accept the lot if $r(n) \leqslant r_A(n)$

2. Reject the lot if $r(n) \geqslant r_R(n)$

3. Continue testing otherwise.

Eqs. 3-12 and 3-13 were adapted from Chapter VIII of Ref. 23; Part II of Ref. 23 is an excellent discussion of sampling plans for variables and attributes. The probabilities of acceptance $P_a(R)$ have 5 special qualities:

$$\left.\begin{aligned} P_a(1) &= 1 \\ P_a(R_0) &= 1-\alpha \\ P_a(1-s) &= a_R/(a_A+a_R) \\ P_a(R_1) &= \beta \\ P_a(0) &= 0 \end{aligned}\right\} \quad (3\text{-}14)$$

These are nominal characteristics; the actual α and β are smaller than shown.

For seq-sampling, the number of items to be tested is not predetermined, but is a random variable whose average is a function of the true reliability. The average sample number ASN_R (number of observations before a decision is reached) for incoming reliability levels of $R = 1, R_0, 1-s, R_1, 0$ is:

$$\left.\begin{aligned} ASN_1 &= a_A/b_2 \\ ASN_{R_0} &= (\bar{\alpha} a_A + \alpha a_R)/(\bar{R}_0 b_1 + R_0 b_2) \\ ASN_{1-s} &= (a_A/b_2)(a_R/b_1) \\ ASN_{R_1} &= (\beta a_A + \bar{\beta} a_R)/(\bar{R}_1 b_1 + R_1 b_2) \\ ASN_0 &= a_R/b_1 \\ &\text{where} \\ \bar{\alpha} &\equiv 1-\alpha \\ \bar{\beta} &\equiv 1-\beta \end{aligned}\right\} \quad (3\text{-}15)$$

Remember that the actual α, β will be appreciably different from those used in Eqs. 3-11 and 3-14. Exact analyses are "available" for some sequential plans, but they are tedious to program for a computer: Ref. 17 has exact analyses for its plans, but they are for the exponential distribution. In general, the tests computed from Eqs. 3-11 and 3-12 can be appreciably truncated without actually exceeding the nominal α, β.

Example No. 24 illustrates the procedure.

3-6.3 EXPONENTIAL ASSUMPTION

The attributes test is similar to the binomial case in terms of test operation and criteria (Ref. 1). The major difference is that R is replaced by the exponential formula $\exp(-\lambda T_M)$

where

T_M = mission time

h = failure rate

The "amount of testing" can be measured by any of the following parameters:

N = number of items on test

r = number of failures

Wt = waiting time before a decision (time elapsed from start of test to the time a decision is reached)

T = total number of accumulated test hours before a decision is reached

If the ARL and URL are specified in terms of failure rate, and if a mission length T_M can be determined for the item, this type of test might be appropriate. This, in turn, will yield

$$R_0(T) = \exp(-\lambda_0 T_M) \tag{3-16a}$$

$$R_1(T) = \exp(-\lambda_1 T_M) \tag{3-16b}$$

the ARL and URL, respectively, for a reliability specification for T_M hours.

The conversion of specified failure rates to probability-of-survival specifications will lead to exactly the same types of tests discussed in pars. 3-6.1 and 3-6.2. However, because of the exponential assumption, the s-expected waiting time before a decision is made can be calculated.

For the 1-sample case, Table 3-5 can be used to determine N and c. It is shown in Ref. 11 that the s-expected waiting time before a decision is reached (as a function of true s-reliability R) is

$$E_R\{Wt_M\} = \sum_{k=1}^{c-1} \binom{N}{k} R^{N-k}\bar{R}^k E_R\{X_{k,N}\} \tag{3-17}$$

where

$$E_R\{X_{k,N}\} = \frac{T_M}{\ln R} \sum_{j=1}^{k} \frac{1}{N-j+1}$$

The term $\sum_{j=1}^{k} \frac{1}{N-j+1}$ is extensively tabulated in Ref. 11 for many sets of k and N.

3-7 EXPONENTIAL PARAMETER, LIFE TESTS

The distribution of failure-times is exponential:

$$R(t) = \exp(-\lambda t) \tag{3-18}$$

where

t = failure time (e.g., time to failure, or time between failures)

$R(t) = Sf\{t\}$

λ = a scale parameter (failure rate); $1/\lambda = \theta = E(t)$

Example No. 24

Design a sequential sample plan so that if the average lot reliability for a 100-hr period is 0.99, there is a 90% probability of acceptance, and if the average lot reliability is 0.95, there is only a 10% probability of acceptance.

Procedure	Example
1. State the parameters of the problem	1. The parameters are $\alpha = 0.10$ $\beta = 0.10$ $R_0 = 0.99,\ \bar{R}_0 = 0.01$ $R_1 = 0.95,\ \bar{R}_1 = 0.05.$
2. Compute a_A, a_R, b_1, b_2 from Eq. 3-11.	2. $a_A = \ln\left(\frac{1 - 0.10}{0.10}\right)$ $= 2.1972$ $a_R = \ln[(1 - 0.10)/0.10]$ $= 2.1972$ $b_1 = \ln(0.05/0.01)$ $= 1.60944$ $b_2 = \ln(0.95/0.90)$ $= 0.05407$ $b_1 + b_2 = 1.6635.$
3. Compute s, h_A, h_R from Eq. 3-12.	3. $s = \frac{0.05407}{1.6094 + 0.05407}$ $= 0.03250$ $h_A = 2.1972/1.6635$ $= 1.3208$ $h_R = 2.1972/1.6635$ $= 1.3208.$

Example No. 24 (Cont'd)

4. Write the equations for the accept and reject lines. Use Eqs. 3-13a and 3-13b.

4. The accept line is

$$r_A = -1.3208 + 0.03250n$$

The reject line is

$$r_R = +1.3208 + 0.03250n.$$

5. Compute the 5 nominal probabilities of acceptance from Eq. 3-14.

5. $P_a(1) = 1$

$$P_a(0.99) = 1 - 10\% = 90\%$$

$$P_a(0.9675) = \frac{2.1972}{2.1972 + 2.1972} = 50\%$$

$$P_a(0.95) = 10\%$$

$$P_a(0) = 0\%.$$

6. Compute the average number of items tested before a decision, for R values in Eq. 3-15.

6. $\mathrm{ASN}_{1.0} = \dfrac{2.1972}{0.05407} = 40.64$

$$\mathrm{ASN}_{0.99} = \frac{(0.90 \times 2.1972) + (0.10 \times 2.1972)}{(0.01 \times 1.6094) + (0.99 \times 0.05407)} = 31.56$$

$$\mathrm{ASN}_{0.9675} = \frac{2.1972}{0.05407} \times \frac{2.1972}{1.6094} = 55.48$$

$$\mathrm{ASN}_{0.95} = \frac{(0.10 \times 2.1972) + (0.90 \times 2.1972)}{(0.05 \times 1.6094) + (0.95 \times 0.05407)} = 16.67$$

$$\mathrm{ASN}_0 = \frac{2.1972}{1.6094} = 1.37 \rightarrow 2.$$

No truncation lines are calculated for this example.

For convenience in the algebra, the tests are discussed in terms of A; the translation to 8 is easily made. Test specifications are presumed to be in the form of a λ_0 for *ARL,* and λ_1 for the *URL.* Specifications given in terms of *R* or 8 can easily be converted to A. Only 2 types of tests are in common use: I-sample and seq-sample.

The concept of total-test-time is important. It is the cumulative operating time of all units regardless of any censoring or of staggered starting times or of replacements. It is to be distinguished from clock (calendar) time.

Replacement affects clock-time, but not total-test-time. The number of test-stations might be limited, or the number of available equipments to go on test might be limited. The test statistic is always total-test-time, but the relationship among number of equipments for test, number of test-stations, and clock-time for the test depends on the initial number on test, whether test-stations are kept occupied (replacement), and other physical strategies. None of these affects the final decision, they affect only the clock-time at which the final decision can be made.

3-7.1 I-SAMPLE

Virtually all I-sample plans are truncated (curtailed), i.e., the test is stopped when a decision boundary is reached; see Fig. **3-2,** with *n* replaced by total-test-time. The average "clock-time to decision" is shorter if more test stations are used. Table 3-7 shows how the average "clock-time to decision" depends on the number of test stations in the nonreplacement case.

The savings in time can be put in quantitative terms, as Table 3-7 shows. To derive Table **3-7,** define $E\{t_{r,s}\}$, the average waiting time to observe the first r failures from s test-stations (no replacement), $s \geqslant r$.

$$\lambda E\{t_{r,s}\} = s \sum_{j=1}^{r} \frac{1}{s - j + 1} \qquad (3\text{-}19)$$

The entries in Table 3-7 are values of the ratio $E\{t_{r,s}\}/E\{t_{r,r}\}$ which quantitatively measures the time saved, and is independent of A.

In many cases, the $(s - r)$ units that have not failed will still be serviceable. If the failure distributions of these units are actually exponential, the survivors will be as good as new. Even if the survivors deteriorate enough to render them unfit for further service, the appreciable savings in time may be worth the cost of the additional units.

Example Nos. **25** and *26* illustrate the procedure.

TABLE 3-7

RATIO OF *s*-EXPECTED WAITING TIMES TO OBSERVE FAILURE *r* IF THERE ARE s TEST-STATIONS

	s							
r	1	2	3	4	5	10	15	20
1	1	0.50	0.33	0.25	0.20	0.10	0.067	0.050
2	...	1	0.56	0.39	0.30	0.14	0.092	0.068
3	...	...	1	0.59	0.43	0.18	0.12	0.087
4	...	...	...	1	0.62	0.23	0.14	0.104
5	...	...	...	...	1	0.28	0.18	0.125
10	...	...	...	...	...	1	0.35	0.23

Example No. 25

Compare the "average clock-time for a test that requires that all 10 units (10 test stations) fail" with the "average clock-time for a test in which 10 units of 20 fail (20 test stations).

Procedure	Example
1. State the values of r and s.	1. $r = 10$, $s = 20$.
2. Determine the ratio of $E\{t_{r,s}\}$ to $E\{t_{r,r}\}$ from Table 3-7.	2. From Table 3-7, ratio = 0.23.

This result clearly indicates the substantial savings in clock-time (on the average) which can result from using more than the minimum number of test stations (in the nonreplacement case).

Example No. 26

Find the 1-sample plan that gives $\alpha = 1\%$, $\beta = 2.5\%$, $\gamma = 2$.

Procedure	Example
1. Use Table V in Ref. 18, or equivalent chi-square tables. Use Eq. 3-21.	1. Use the 2 columns $csqf(\chi^2; \nu) = 1\%$ $csqf(\chi^2; \nu) = 100\% - 2.5\% = 97.5\%$.
2. By trial and error, find the ν such that Eq. 3-23 is satisfied. Use only even ν since $r^* = \nu/2$ is an integer. The underline indicates a good answer.	2. $\dfrac{\chi^2_{97.5\%,\nu}}{\chi^2_{1\%,\nu}} = 2$ $\nu = 30$: 47.0/15.0 = 3.13 $\nu = 40$: 59.3/22.2 = 2.67 $\nu = 60$: 83.3/37.5 = 2.22 $\nu = 74$: 99.7/48.7 = 2.05 $\underline{\nu = 80}$: 106.6/53.5 = 1.993 $\nu = 78$: 104.3/51.9 = 2.01.0.
3. Calculate r^*. Rechoose α, β, γ to fit.	3. $r^* = 80/2 = 40$ choose $\gamma = 1.993$, then $\alpha = 1\%$, $\beta = 2.5\%$ still.
4. Use Eqs. 3-21 and 3-22 to find T^*.	4. $\chi^2_{1\%,80} = 2\lambda_0 T^* = 53.5$ $\chi^2_{97.5,80} = 2\lambda_1 T^* = 106.6$ $T^* = 26.75/\lambda_0 = 53.3/\lambda_1$ $= 26.75\theta_0 = 53.3\theta_1$.

Twenty seven to 53 (from last line of step 4) is an outlandish maximum number of failures for any large equipments; so the α, β, γ cannot all be this small. The time truncation, about $27\theta_0$, is also outlandish for any long-life item.

The decision boundaries for the 1-sample test are shown in Fig. 3-4(A). When the path of the test touches or crosses a boundary ($r \geq r^*$, or $T = T^*$) the appropriate decision is made. The *OC* for the plan does not depend on the number of test stations nor on the replacement policy.

$$\left.\begin{aligned} P_a(\lambda) &= \sum_{r=1}^{r^*-1} e^{-\mu}\mu^r/r! \\ &= poif(r^* - 1;\mu) \\ &= csqfc(2\mu; 2r^*) \end{aligned}\right\} \quad (3\text{-}20)$$

P_a = probability of an "accept" decision

r^* = number of failures for rejection

μ $\equiv AT^*$, mean number of failures for failure rate A in time T^*

λ = failure rate

T^* = total-test-time for acceptance

$csqfc\ (\chi^2; \nu)$ = complement of the χ^2 *Cdf* with ν degrees of freedom

Test plans are available in Table 3-8, Ref. 9, Ref. 17, and elsewhere. **Eq.** 3-20 is the basis for the plans. By definition of α and β, Eqs. 3-21a and 3-21b are true.

$$\alpha = csqf(2\lambda_0 T^*; 2r^*) \quad (3\text{-}21a)$$

$$1 - \beta = csqf(2\lambda_1 T^*; 2r^*) \quad (3\text{-}21b)$$

Define $\chi^2_{P,\nu}$ such that

$$P = csqf(\chi^2_{P,\nu}; \nu) \ . \quad (3\text{-}22)$$

Then **Eq.** 3-21 becomes

$$\frac{\chi^2_{1-\beta,2r*}}{\chi^2_{\alpha,2r*}} = \frac{\lambda_1}{\lambda_0} = \gamma \quad (3\text{-}23)$$

which relates r^*, a, β, γ. If r^* and 2 **of** the others are specified, the third can be determined exactly. Since r^* must be an integer, it is not possible to specify α, β, γ and determine r^* exactly. The approximate value of r^* can be determined, then the other 3 adjusted to give "reasonable" values. See Example No. 26.

To find the *s*-expected clock-time of the test, it is convenient to calculate $E\{r\}$. See Fig. 3-4(A). If the test path hits the line A, the probability of a particular r is just the Poisson formula with $\mu = AT^*$, and $0 \leq r < r^*$. If the test path hits line R, it is equivalent to hitting the upward extension of A for $r^* \leq r$; the number of failures is r^*, but the probability is the Poisson formula for r, with $\mu = AT^*$. Therefore the average r is

$$\left.\begin{aligned} E\{r; \lambda, T^*, r^*\} &= \sum_{r=0}^{r^*-1} r\left(\frac{e^{-\mu}\mu^r}{r!}\right) \\ &\quad + r^* \sum_{\lambda=r^*}^{\infty} \frac{e^{-\mu}\mu^r}{r!} \\ &= \sum_{r=0}^{r^*-2} \frac{e^{-\mu}\mu^r}{r!} \\ &\quad + r^*\left(1 - \sum_{\lambda=0}^{r^*-1} \frac{e^{-\mu}\mu^r}{r!}\right) \\ &= \mu\ csqfc(2\mu; 2r^* - 2) \\ &\quad + r^* csqf(2\mu; 2r^*) \end{aligned}\right\} \quad (3\text{-}24)$$

This formula is also derived in Ref. 11. It is straightforward to **show** that

$$AE(T; A, T^* r^*\} = E\{r; A, T^*, r^*\} \quad (3\text{-}25)$$

and that (for replacement)

$$sE\{\text{clock-time}; A, T^*, r^*\} = E\{T; A, T^*, r^*\} \quad \textbf{(3-26)}$$

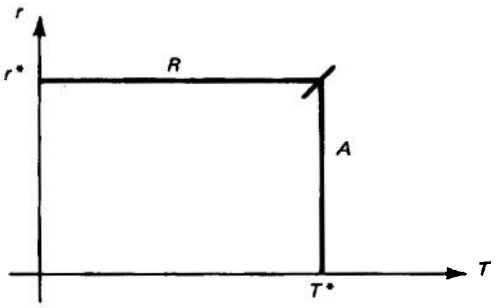
r = number of failures so far

T = total test time so far

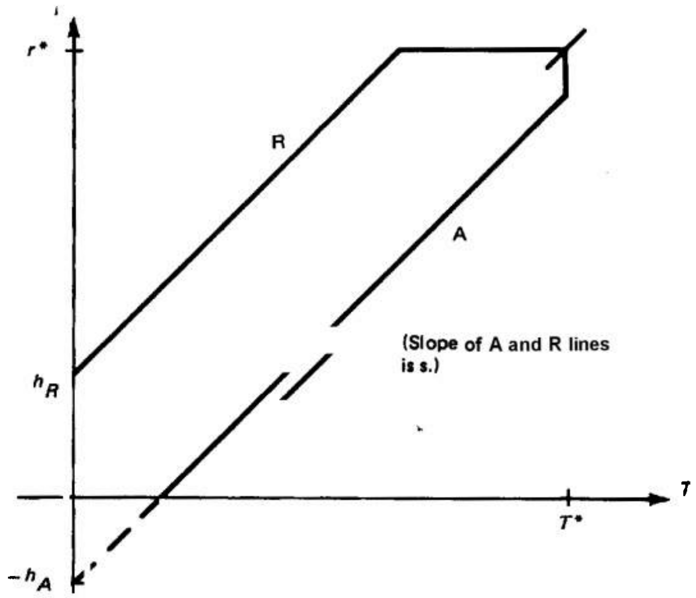
r^*, T^* = boundaries for r and T

A = "Accept" portion of boundary

R = "Reject" portion of boundary



(A) 1-sample plan



(B) Seq-sample plan 2-way decision

Figure 3-4. Tests for the Exponential Parameter where s = number of test stations

$\mu \equiv \lambda T$

$E\{r, \lambda_0\}$ and $E\{r, \lambda_1\}$, for various test plans, are given in Table 3-8.

3-7.2 SEQ-SAMPLE

Consult pars. 3-5 and 3-6 for a general discussion of seq-sample tests and references for further reading. Fig. 3-4(B) shows the decision boundary for a typical test. The equations for the lines are similar to those in par. 3-6 for the binomial parameter (adapted from Ref. 13).

$$\left.\begin{aligned}
a_A &= \ln\left(\frac{1-\alpha}{\beta}\right) \\
a_R &= \ln\left(\frac{1-\beta}{\alpha}\right) \\
b_1 &= \lambda_1 - \lambda_0; \\
b_{1,0} &= \frac{\lambda_1}{\lambda_0} - 1 = b_1/\lambda_0; \\
b_{1,1} &= 1 - \frac{\lambda_0}{\lambda_1} = b_1/\lambda_1 \\
b_2 &= \ln \lambda_1 - \ln \lambda_0 = \ln(\lambda_1/\lambda_0)
\end{aligned}\right\} \quad (3\text{-}27)$$

$$\left.\begin{aligned}
h_A &= a_A/b_2 && \text{(intercept)} \\
h_R &= a_R/b_2 && \text{(intercept)} \\
s &= b_1/b_2;\ s_0 = b_{1,0}/b_2; \\
s_1 &= b_{1,1}/b_2 && \text{(slope)}
\end{aligned}\right\} \quad (3\text{-}28)$$

$$\begin{aligned}
r_A &= -h_A + sT = -h_A + s_0 \times (\lambda_0 T) \\
&= -h_A + s_1 \times (\lambda_1 T)
\end{aligned} \quad (3\text{-}29a)$$

$$\begin{aligned}
r_R &= h_R + sT = h_R + s_0 \times (\lambda_0 T) \\
&= h_R + s_1 \times (\lambda_1 T)
\end{aligned} \quad (3\text{-}29b)$$

where

α = producer risk; probability of rejecting a lot with $\lambda = \lambda_0$, $\lambda_0 < \lambda_1$

β = consumer risk; probability of accepting a lot with $\lambda = \lambda_1$, $\lambda_0 < \lambda$

T = total-test-time

It is often more convenient to use a normalized total-test-time $\lambda_0 T$ or $\lambda_1 T$, rather than T, itself.

The minimum accept time is for $r_A = 0$, and is (from Eq. 3-29a)

$$\lambda_0 T_{min} = h_A/s_0 = a_A/b_{1,0} \quad (3\text{-}30a)$$

$$\lambda_1 T_{min} = h_A/s_1 = a_A/b_{1,1} \quad (3\text{-}30b)$$

The minimum number of failures to reject is for $T = 0$, and is (from Eq. 3-29b)

$$r_{R,min} = h_R = a_R/b_2 \quad (3\text{-}31)$$

The computation of the exact α,β with truncation is done with special computer programs. The basic algorithm is simple enough, but the calculations are horrendous for large tests. Ref. 17 shows the exact α, β for the truncated plans. In practice, the truncation of a plan as calculated from Eqs. 3-27, 3-28, and 3-29 can be severe without having the actual α,β exceed the nominal ones.

The usual difficulties occur with this test plan as with any such plan. If λ_1 and λ_0 are close together, and α and β are small, the test-time for λ near s can be quite long. Ref. 15 discusses some of these difficulties; they are far from being resolved.

It is usually convenient to work with $\alpha = \beta$. It makes little difference how they are chosen if the discrimination ratio is available for adjusting. After all, we are just picking 2 points on the *OC* curve and they might as well be convenient ones. Then s_0 and s_1 can be written in terms of the discrimination ratio. The subscripts A and R on a and h can be dropped. Eqs. 3-27,

TABLE 38

TEST PARAMETERS AND s-EXPECTED NUMBER OF FAILURES FOR VARIOUS 1-SAMPLE AND SEQSAMPLE LIFE TESTS (ADAPTED FROM Ref. 1)

			1-sample plans				seq-sample plans		
					$E\{r/\lambda\}$			$E\{r/\lambda\}$	
$\gamma = \frac{\lambda_1}{\lambda_0}$	α	β	Rejection Number, r^*	$\chi^2/2$ $1-\alpha, 2r$	$\lambda = \lambda_0$	$A = \lambda_1$	Truncation Number, r^*	$\lambda = \lambda_0$	$\lambda = \lambda_1$
1.5	0.05	0.05	67	54.13	54.0	66.8	201	28.0	36.7
	0.05	0.10	55	43.40	40.5	54.6	165	21.1	32.9
	0.05	0.25	35	25.87	24.0	34.0	105	12.0	23.5
	0.10	0.05	52	43.00	37.6	51.8	156	25.1	27.6
	0.10	0.10	41	33.04	32.8	40.7	123	18.6	24.4
	0.10	0.25	25	18.84	18.7	24.2	75	10.1	16.5
	0.25	0.05	32	28.02	27.3	31.9	96	18.0	15.7
	0.25	0.10	23	19.61	9.0	22.7	69	12.6	13.2
	0.25	0.25	12	9.52	9.1	11.4	36	5.8	7.6
2	0.05	0.05	23	15.72	5.6	22.9	69	8.6	13.7
	0.05	0.10	19	12.44	2.4	18.8	57	6.5	12.3
	0.05	0.25	13	7.69	7.6	12.4	39	3.7	8.8
	0.10	0.05	18	12.82	2.7	17.9	54	7.7	10.3
	0.10	0.10	15	10.30	10.2	14.8	45	5.7	9.1
	0.10	0.25	9	5.43	5.3	8.5	27	3.1	6.2
	0.25	0.05	11	8.62	8.2	10.9	33	5.5	5.9
	0.25	0.10	8	5.96	5.6	7.8	24	3.9	4.9
	0.25	0.25	5	3.37	3.2	4.7	15	1.8	2.8
3	0.05	0.05	10	5.43	5.4	9.9	30	2.9	6.1
	0.05	0.10	8	3.98	3.9	7.8	24	2.2	5.5
	0.05	0.25	6	2.61	2.6	5.6	18	1.3	3.9
	0.10	0.05	8	4.66	4.6	7.9	24	2.6	4.6
	0.10	0.10	6	3.15	3.1	5.9	18	2.0	4.1
	0.10	0.25	4	1.74	1.7	3.6	12	1.1	2.8
	0.25	0.05	5	3.37	3.2	5.0	15	1.9	2.6
	0.25	0.10	4	2.54	2.4	3.9	12	1.3	2.2
	0.25	0.25	2	0.96	0.86	1.7	6	0.61	1.3
5	0.05	0.05	5	1.97	1.9	5.0	15	1.1	3.3
	0.05	0.10	4	1.37	1.4	3.9	12	0.83	2.9
	0.05	0.25	3	0.82	0.81	2.7	9	0.47	2.1
	0.10	0.05	4	1.74	1.7	4.0	12	0.99	2.5
	0.10	0.10	3	1.10	1.1	2.9	9	0.73	2.2
	0.10	0.25	3	1.10	1.1	2.9	9	0.40	1.5
	0.25	0.05	2	0.96	0.86	1.9	6	0.71	1.4
	0.25	0.10	2	0.96	0.86	1.9	6	0.50	1.2
	0.25	0.25	1	0.29	0.26	0.8	3	0.23	0.68

3-28, and *3-29* become

$$\left.\begin{aligned}&\alpha = \beta\\&\gamma = \lambda_1/\lambda_0\\&a = \ln\left(\frac{1-\alpha}{\beta}\right) = \ln\left(\frac{1-\beta}{\alpha}\right)\\&b_{1,0} = \gamma - 1,\ b_{1,1} = 1 - 1/\gamma\\&b_2 = \ln\gamma\end{aligned}\right\} \quad (3\text{-}32)$$

$$\left.\begin{aligned}&h = a/b_2\\&s_0 = (\gamma - 1)/(\ln\gamma),\\&s_1 = \left(1 - \frac{1}{\gamma}\right)\Big/(\ln\gamma) = s_0/\gamma\end{aligned}\right\} \quad (3\text{-}33)$$

$$\left.\begin{aligned}r_A &= -h + s_0 \times (\lambda T)\\&= -h + s_1 \times (\lambda_1 T)\\r_R &= h + s_0 \times (\lambda T)\\&= h + s_1 \times (\lambda_1 T)\end{aligned}\right\} \quad (3\text{-}34)$$

It is now worthwhile tabulating b_2, s_0, s_1 as functions of λ; see Table 3-9.

Example Nos. 27 and 28 illustrate the procedure.

TABLE 3-9

FACTORS FOR SEQ-SAMPLE TESTS

See Eqs. 3-32, 3-33, and 3-34 in the text.

γ	b_2	s_0	s_1
1.25	0.223	1.120	0.896
1.50	0.405	1.233	0.822
2.0	0.693	1.443	0.721
2.5	0.916	1.637	0.655
3.0	1.099	1.820	0.607
3.5	1.253	1.996	0.570
4.0	1.386	2.164	0.541
5.0	1.609	2.485	0.497

3-8 *s*-NORMAL PARAMETER, MEAN

The standard deviation is presumed to be known exactly, and the mean is the parameter upon which acceptance rests. The 1-sample plans are well known in the quality control field. Seqsample plans are feasible, and some forms of the plans are given in this paragraph. No truncation data are readily available. The equations are adapted from Ref. *23,* Chap. XV.

Notation follows:

μ = actual mean

μ_H = higher mean

μ_L = lower mean

$\Delta\mu$ = $(\mu_H - \mu_L)/2$

σ = known standard deviation

α_H, α_L = probability of incorrect decision when μ is μ_H or μ_L

x = random sample from s-normal population with mean μ and standard deviation σ

Z_L, Z_H = $(x - \mu_L)/\sigma$ or $(x - \mu_H)/\sigma$

γ = $(\Delta\mu)/\sigma$, discrimination ratio

Σ_i = sum over all measurements so far

n = number of individual items tested so far

The equations for the choose "μ is Low" or "μ is High" are (when μ_L is used as a reference)

Example No. 27

A large digital computer is designed as the heart of a battlefield intelligence and fire direction system. Since only three systems are available for test, and test time is limited to 2000 hr each, it is decided to use a sequential test. Investigate a sequential test which demonstrates a minimum acceptable mean life of 1000 hr for a consumer and producer risk of 0.1.

Procedure	Example
1. State the given test parameters and choose a set of suitable equations.	1. $\alpha = \beta = 10\%$ $\lambda_1 = 1/1000\text{-hr}$ Use Eqs. 332, 333, 3-34 with $\lambda_1 T$.
2. Calculate a from Eq. 332.	2. $a = \ln\left(\frac{1-0.1}{0.1}\right) = 2.197.$
3. Construct a table which shows s_1, h, and λ_0 as a function of γ. Use Table 3-9 and Eqs. 3-32, 3-33, 3-34, and 3-30.	3. (see table below)

γ	s_1	h	$\lambda_1 T_{min}$
1.5	0.822	5.419	6.59
2.0	0.721	3.170	4.40
3.0	0.607	2.000	3.30
5.0	0.497	1.365	2.75

Now look at the table just constructed. If the supplier is willing to take a 10% chance of not passing the test (a large chance to take on a big contract), he must, for example, make the item 3 times as good as is required ($\gamma = 3$) and test it for "$3.30/\lambda_1 = 3300$ hr ≈ 20 weeks" of total test time at a minimum (presumes no failures). For every failure, he must test for another "$1/(s_1\lambda_1) = 1650$ hr ≈ 10 weeks" if he is to pass the test. If he chooses to strive only for twice as good as necessary ($\gamma = 2$), then he must test for a minimum of "$4.40/\lambda_1 = 4400$ hr ≈ 26 weeks" of total test time; for every failure he must test for another "1387 hr ≈ 8 weeks." Most suppliers don't want to take a 10% chance of failing such an important test, especially if their equipment is 2 to 3 times as good as required. The additional requirement that total test time be no more than "3 × 2000 hr = 6000 hr" hampers things even more; 6000 hr corresponds to a $\lambda_1 T$ of 6, which means that the $\gamma = 1.5$ option is not even open to him.

Example No. 28

Same data as Example No. 27 but use Ref. 17 to solve the problem.

Procedure	Example
1. Pick several plans that have $\alpha = \beta = 10\%$.	1. In version B, Rev. July 1969, there are 2 plans which might be applicable: V and VI.
2. Convert the time axis from $\lambda_0 T$ to $\lambda_1 T$, because this example is given in terms of λ_1.	2. Multiply times in Plan V by $\gamma = 3$, and in Plan VI by $\gamma = 5$.
3. Find $\lambda_1 T_{max}$ (truncation).	3. Plan V, $\lambda_1 T_{max} = 10.35$ Plan VI, $\lambda_1 T_{max} = 6.25$.
4. Choose a plan.	4. Because of the "T_{max} = 6000 hr" ($\lambda_1 T = 6$) constraint, Plan VI is preferred over Plan V. It has $\gamma = 5$, and actual α, β of about 13%, (see p. 60, Ref. 17).
5. Check the OC curve and s-expected test time curve.	5. (See p. 68, Ref. 17). To get a rejection probability of 4% requires $\lambda/\lambda_1 \approx 10$, and of 2% requires $\lambda/\lambda_1 \approx 15$. The s-expected test time for very good equipment is about "$3/\lambda_1$ = 3000 hr."

So the test plan that fits the problem constraints requires, in essence, that the equipment be 10 times as good as required, just to keep from having an unreasonable chance of failing the test when the equipment is much better than needed.

This example illustrates the paradox brought on by the exponential distribution and expensive tests.

$$\left.\begin{array}{l}\text{Choose "}\mu\text{ is High":}\\ \Sigma_i Z_{L,i} = (a_H/\gamma) + \gamma n\\ \text{Choose "}\mu\text{ is Low":}\\ \Sigma_i Z_{L,i} = -(a_L/\gamma) + \gamma n\end{array}\right\} \quad (3\text{-}35)$$

where

$$\left.\begin{array}{l} a_H \equiv \ln\left(\dfrac{1-\alpha_H}{\alpha_L}\right)\\ a_L \equiv \ln\left(\dfrac{1-\alpha_L}{\alpha_H}\right)\end{array}\right\} \quad (3\text{-}36)$$

Where μ_H is used as a reference, Eqs. 3-35 become

$$\left.\begin{array}{l}\text{Choose "}\mu\text{ is High":}\\ \Sigma_i Z_{H,i} = (a_H/\gamma) - \gamma n\\ \text{Choose "}\mu\text{ is Low":}\\ \Sigma_i Z_{H,i} = -(a_L/\gamma) - \gamma n\end{array}\right\} \quad (3\text{-}37)$$

For use in an actual situation, the Z's would be converted to x's for ease in use. Eqs. 3-35 and 3-37 are helpful in visualizing what happens when the discrimination ratio is changed.

3-9 BAYESIAN STATISTICS

A good approach to Bayesian statistics is given in Ref. 25; an idea of its breadth of application is found in Ref. 26; and an easy-to-read discussion of its controversial nature is propounded in Ref. 27. The mathematics of Bayesian statistics is not controversial at all. It is the interpretation of those equations concerning knowledge as "prior", and what constitutes reasonable "prior" knowledge which is the source and mainstay of the controversy.

Engineers, as well as most other people, tend to confuse what they hope and want to be true, with what they really expect is true. If one is to use Bayesian statistics, he must evaluate the consequences of his assumptions very thoroughly, before running any tests. (Portions of this paragraph are reprinted from Ref. 28, with permission.)

3-9.1 PROBABILITY AND BAYES PROBABILITY

Probability is a mathematical concept used in connection with random events, i.e., those events whose occurrence is uncertain enough that the uncertainty is of concern to us. One of the most popular uses of probability is in games of chance; we speak of the odds or percentages. For example, in a pair of honest dice the probability of throwing "snake eyes" is 2-7/9%; the probability of rolling 7's is 16-2/3%. In these uses of probability, it can be shown that the probability is associated with the "long run" percentages.

Another popular use of probability is as degree-of-belief. We speak of the probability of winning a case at law or of getting a promotion. If a person is prudent, his degree-of-belief is the same as the "long run" percentage, when that percentage is known. But "long run" percentages usually are associated with conceptual models such as unbiased coins and honest dice. It is degree-of-belief as to whether the coin is in fact reasonably unbiased, or the dice are actually honest.

The term Bayes probability has become associated with degree-of-belief and it will be used in that sense in this paragraph. (Not everyone uses it that way—the language is in a state of flux.) The Bayes formula provides a means of converting the degree-of-belief we had "before a test was run" to the degree-of-belief we have afterwards; degree-of-belief is not a static thing, we change it whenever we get more evidence. Bayes formula provides the mathematics by which a rational person has his degree-of-belief changed by evidence.

3-9.2 SIMPLE ILLUSTRATION

To use the Bayes procedure, we must first state what our beliefs are about every possibility in our conceptual model. Suppose we wish to take a stand on whether a coin is honest or not, and all we can see of it are the results of legitimate flips. Row 1 in Table 3-10 shows the 3 conditions we presume are possible in our conceptual model; the coin is either 2-tailed, fair, or 2-headed. Row **2** shows our degree-of-belief before any tests are run; we are **99%** sure the coin is fair, and we suspect 1/2% each that the coin is 2-tailed or 2-headed. Let each test involve seeing the results of **3** legitimate flips of the coin. Run Test A and suppose the results are 3-heads, 0-tails. Now, separately for each condition, calculate the likelihood of getting the test result if that condition is true. Obviously, under condition #1 (2-tailed coin) for example, the results of Test A would be impossible; so we put a zero there. By means of Bayes formula, our "after test A" degree-of-belief is calculated. Reasonably enough, we now suspect more strongly that the coin is not a fair one, but we no longer suspect that it is 2-tailed. After the results of Test B are in, we are even more suspicious about the coin's being 2-headed. But Test *C* clears things up; since we have observed at least 1 tail and 1 head, the coin can be neither 2-headed nor 2-tailed. Therefore it must be fair, according to the conceptual model we set up. The "after test" degree-of-belief often serves as the "before test" degree-of-belief for a subsequent test, as in this illustration.

3-9.3 BAYES FORMULAS, DISCRETE RANDOM VARIABLES

Bayes formula is

$$Pr\{E_i \mid H\} \propto Pr\{E_i\} Pr\{H \mid E_i\} \tag{3-38}$$

$$\Sigma_i Pr\{E_i \mid H\} = 1 \tag{3-39}$$

TABLE 3-10

HONEST COIN?

	Condition	No. 1	No. 2	No. 3
(1)	**probability of heads**	**0**	**1/2**	**1**
(2)	degree-of-belief	1/2%	99%	1/2%
(3)	likelihood of test A results 3-heads, 0-tails	0	1/8	1
(4)	degree-of-belief	0%	96%	4%
(5)	likelihood of test B results 3-heads, 0-tails	0	1/8	1
(6)	degree-of-belief	0%	76%	24%
(7)	likelihood of test C results 1-heads, 2-tails	0	3/8	1
(8)	degree-of-belief	0%	100%	0%

where

E_i – subevent i for E. The event space is partitioned (exhaustive and mutually exclusive subevents) into subevents.

H – event: new test results

$Pr\{E_i\}$ = prior probabilities assigned to the E_i

$Pr\{H|E_i\}$ = likelihood of getting the new results, given that E_i were in fact true.

$Pr\{E_i|H\}$ = new probabilities assigned to the E_i, after seeing the test results.

Σ_i = implies sum over all i

In the illustration in par. **3-9.2** (see Table 3-10), rows **4**, *6*, **8** were obtained from **Eqs. 3-38** and **3-39** as shown in Example No. 29.

3-9.4 PRINCIPLES FOR APPLICATION

Apply the following principles:

Example No. 29

~~Procedure~~	~~Example~~
1. State prior probabilities.	1. $Pr\{E_1\} = 0.005$, $Pr\{E_2\} = 0.99$, $Pr\{E_3\} = 0.005$.
2. Experimental result H (test A) was 3-heads, 0-tails. Calculate $Pr\{H \mid E_i\}$: row 3 in Table 3-10.	2. E_1 = 2-tailed coin, $Pr\{H \mid E_1\} = 0$; E_2 = honest coin, $Pr\{H \mid E_2\} = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$; E_3 = 2-headed coin, $Pr\{H \mid E_3\} = 1$.
3. Calculate $Pr\{E_i \mid H\}$. Use Eq. 3-38 first.	3. $Pr\{E_1 \mid H\} \propto 0.0050 \times 0 = 0$ $Pr\{E_2 \mid H\} \propto 0.990 \times (1/8) = 0.1238$ $Pr\{E_3 \mid H\} \propto 0.0050 \times = \underline{0.0050}$ Total 0.1288
Use Eq. 3-39 next. Divide by the normalizing factor (the total).	$Pr\{E_1 \mid H\} = 0/0.1288 = 0$ $Pr\{E_2 \mid H\} = 0.1238/0.1288 = 0.9612$ $Pr\{E_3 \mid H\} = 0.0050/0.1288 = 0.0388$
4. New experimental result H (test B) was again 3-heads, 0-tails. Calculate $Pr\{H \mid E_i\}$: row 5 in Table 3-10.	4. $Pr\{H \mid E_1\} = 0$ $Pr\{H \mid E_2\} = (\frac{1}{2})^3 =$ [redacted] $Pr\{H \mid E_3\} = 1$.
5. Calculate $Pr\{E_i \mid H\}$. Use Eq. 3-38 first. The probabilities after test A are now the prior probabilities for test B.	5. $Pr\{E_1 \mid H\} \propto 0 \times 0 = 0$ $Pr\{E_2 \mid H\} \propto 0.9612 \times (1/8) = 0.1202$ $Pr\{E_3 \mid H\} \propto 0.0388 \times 1 = \underline{0.0388}$ Total 0.1590
Use Eq. 3-39 next. Divide by the normalizing factor (the total).	$Pr\{E_1 \mid H\} = 0$ $Pr\{E_2 \mid H\} = 0.1202/0.1590 = 0.7562$ $Pr\{E_3 \mid H\} = 0.0388/0.1590 = 0.2441$

Example No, 29 (Cont'd)

6. New experimental result H (test C) was 1-head, 2-tails. Calculate $Pr\{H \mid E_i\}$: row 7 in Table 3-10.

6. $Pr\{H \mid E_1\} = 0$

$Pr\{H \mid E_2\} = 3 \times (\frac{1}{2}) \times (\frac{1}{2})^2 = \frac{3}{8}$

$Pr\{H \mid E_3\} = 0.$

7. Calculate $Pr\{E_i \mid H\}$. Use **Eq. 3-38** first. The probabilities after test B are now the prior probabilities for test C.

7. $Pr\{E_1 \mid H\} \propto 0 \times 0 = 0$

$Pr\{E_2 \mid H\} \propto 0.7562 \times (3/8) = 0.2836$

$Pr\{E_3 \mid H\} \propto 0.2441 \times 0 = \underline{0}$

Total 0.2836

Use **Eq. 3-39** next as in step 5.

$Pr\{E_1 \mid H\} = 0/02836 = 0$

$Pr\{E_2 \mid H\} = 0.2836/0.2836 = 1$

$Pr\{E_3 \mid H\} = 0/0.2836 = 0.$

This simple illustration shows how Bayes formula can be applied repeatedly to a problem as more test data become available.

Once the probability of any E, is zero, it can never again be anything but zero. Thus, it is important to allow at least some initial prior probability for any possible outcome.

(1) Each possible condition of the unknown (in the conceptual model) must be specified. Seriously consider using discrete values to represent conditions, because we are used to thinking that way. Continuous values. e.g., between 0 and 1, can be used, but then step 2 is often very deceptive because we are not used to thinking in terms of the probability density functions that are then needed; it's like trying to guess which shell the pea is under.

(2) Assign a prior degree-of-belief to each of the conditions, Leaving out a condition is equivalent to assigning zero degree-of-belief to it. Once zero degree-of-belief has been assigned to a condition, or calculated for it after a test, the degree-of-belief for that condition remains zero forever after. *So* we must include (at least approximately) all physically possible conditions. **As** we shall see in the example in par. **3-9.5,** we must not assign too low a degree-of-belief to the unlikely regions. **A** good way of handling the assignment is to pretend that we are willing to bet money on the outcomes at the odds we have implicitly specified and that the other person can choose whichever side of the bet he wishes.

Exercise your model with hypothetical test results; see if your afterwards degrees-of-belief correspond to the calculated ones. If not, go back and change your prior degrees-of-belief.

(3) Run a test which is intended to shed some light on the unknown. Preferably, the likelihood of the test results should be quite different for each condition.

(4) For each condition, calculate the likelihood of the test results. If the test was chosen well, the likelihood will be quite different for each condition.

(5) Use Bayes formula to calculate the "after test" degree-of-belief.

(6) If the degree-of-belief is not sharp enough (high for only a few close-together conditions, and low for the rest) consider running more tests.

When a single point estimate is desired (instead of the distribution), we can choose the most probable value, the average value, the median value, or some other that we prefer. If these reasonable choices are close together, it makes little difference which we choose. If they are far apart, we're in trouble anyway—we need more tests.

Step 2 is deceptively simple looking. In practice it is easy enough to assign a seemingly reasonable degree-of-belief to each condition. But we may well not like the "after test" degree-of-belief we're supposed to have. **A** story illustrates the point. **A** man kept asserting to his friends that he was dead. The friends finally persuaded him to see a medical doctor. The doctor hit upon the idea of a test; said he, "Dead men don't bleed do they?" The man readily agreed with the assumption that dead men do not bleed and that the test would be reasonable. Thereupon the doctor stuck the man's thumb which bled profusely. The man looked at the test results and exclaimed, "By golly, dead men do bleed!"

Step 2 is the place where most Bayesian disasters occur. The utter simplicity of the formulas belies the skill and hard work needed to apply prior probabilities properly. What you think you believe, and what you actually believe, after serious analysis, are often very, very different. The literature abounds with examples of engineers who tried Bayesian analysis without the skill and hard work.

3-9.5 COMPLEX ILLUSTRATION

Any assignment of prior degree-of-belief should be thoroughly checked by simulating test results before submitting to any real tests. During the simulation, we may wish to change our assignment of prior degree-of-belief.

Suppose a portable power tool is to be tested for insulation breakdown under super-severe conditions (high temperature, high load, high humidity, high vibration). If the insulation breaks down in field use, injury or death could result. We generate row 1 of Table 3-11 by considering what we want to know. If the failure probability is high, we don't really care exactly what it is because it's "back to the drawing board" anyway. The lower the failure probability, the closer we want to know it. In order to keep the example simple, we are supposing that only *6* possible failure probabilities exist, as shown in row 1. Row 2 shows the degree-of-belief of the designer before submitting to this accelerated test.

Now let's simulate. Suppose the test results are 0-pass, 2-fail. Row **4** shows the "after test" degree-of-belief for each condition. No one in his right mind still believes there's an almost 50-50 chance that the failure probability is 0.03. The analysis was correct; so the assumptions were bad. The designer was too optimistic—a very common circumstance. Row ***5*** is a revised "before test" degree-of-belief. It still has a large peak at condition *5,* but the bad conditions have higher degrees-of-belief. Now with **an** hypothesized 0-pass, 2-fail test result, the "after test" degree-of-belief in row **7** seems more reasonable. Let's try the same "before test" degree-of-belief with 2-pass, 0-fail results; as shown in rows 8-10. Row 10 gives the "after test" degree-of-belief, and it too seems reasonable. In practice, you should try many more simulations of test results before coming to a conclusion about your "before test" degree-of-belief.

You may wish to have your problem programmed for a computer; then you can simulate much more extensively and easily

Sooner or later, someone will advise you to use an "ignorance" prior, i.e., a prior that assumes "complete ignorance of the situation". Reject that advice. Proceed by putting down what you think you do believe;

TABLE 3-11

PORTABLE POWER TOOL INSULATION TEST

	Condition	#1	#2	#3	#4	#5	#6
(1)	probability of failing test	0.9	0.5	0.2	0.1	0.03	0.01
(2)	"before test" d-of-b	0.1%	0.1%	0.1%	0.7%	95%	4%
(3)	likelihood of 2-fail, 0-pass	0.81	0.25	0.04	0.01	0.0009	0.0001
(4)	"after test" d-of-b	40%	12%	1.9%	3.9%	42%	0.2%
(5)	"before test" d-of-b	1%	1%	1%	1%	95%	1%
(6)	likelihood of 2-fail, 0-pass	0.81	0.25	0.04	0.01	0.0009	0.0001
(7)	"after test" d-of-b	68%	21%	3.2%	0.8%	7%	0.01%
(8)	"before test" d-of-b	1%	1%	1%	1%	95%	1%
(9)	likelihood of 0-fail, 2-pass	0.01	0.25	0.64	0.81	0.94	0.98
(10)	"after test" d-of-b	0.01%	0.3%	0.7%	0.9%	97%	1.1%

d-of-b = degree-of-belief, i.e., Bayesian probability then check it out by simulation; revise it and check again; repeat the process until you are satisfied that you believe what you've recorded.

Bayesian analysis is tedious and time consuming when it is done right. When it is done wrong (sloppily), it can be very misleading.

3-9.6 BAYES FORMULAS, CONTINUOUS RANDOM VARIABLES

$$pdf\{\phi \mid \hat{\phi}\} \propto pdf\{\phi\}pdf\{\hat{\phi} \mid \phi\} \qquad (3\text{-}40)$$

$$\int_{\phi} pdf\{\phi \mid \hat{\phi}\}d\phi = 1 \qquad (3\text{-}41)$$

where

d) = value of the continuous random variable

$\hat{\phi}$ = estimate of d) made from a new test

$pdf\ \{\phi\}$ = prior pdf assigned to d) (before the test)

$pdf\ \{\hat{\phi}|\phi\}$ = likelihood of getting the test result $\hat{\phi}$ given that d) is the true value

$pdf\ \{\phi|\hat{\phi}\}$ = new pdf assigned to ϕ after seeing the test results

$\int_{\phi}$ implies the integral over the domain of d)

Other forms of the equations are possible for other combinations of discrete and random variables for the unknown variable and the prior information.

3-9.7 CONJUGATE PRIOR DISTRIBUTIONS

When the events and probabilities are in the form of a probability distribution, it is mathematically possible to find a form of prior distribution such that the after-test (posterior) distribution has the same parametric form (from the same family of distributions) as the prior; see Ref. 25 or other Bayesian textbook.

In this case the new distribution is the same as the old, except for different values of the parameters. As an example, for the usual failure rate, one can hypothesize a prior distribution with parameters r and T where r = number of failures and T = total test time. The estimate of A usually used is simply r/T. The new estimate of λ is made by adding the new increments in r and T to the old values—just as if a classical test were being run. But, in Bayesian statistics, the old r and T don't have to be actual test results, they can be numbers that are equivalent to your prior degree-of-belief.

The calculational simplicity of this approach has much appeal, and its use is popularized in books and articles. Before using it, compare it with the discrete method described earlier in this paragraph. Use simulation to find the consequences of your assumptions. But remember, computational simplicity is not the purpose of a Bayesian analysis. The real purpose is to find a reasonable way to put your prior knowledge to work.

Example No. 30 illustrates the procedure.

3-9.8 LIFE-TESTING

Ref. 29 shows how Bayesian techniques can be applied to accept/reject tests. Be very careful in using them. Simulate extensively before applying any in an important situation. The mathematics is deceptively

Example No. 30

Suppose that your prior degree-of-belief about a failure rate is that it is equivalent to having found 4 failures in 2000 hr of testing (λ_{est} = 1/2000-hr). Suppose the test results are 3 failures in 1000 hr. What is your new degree-of-belief? Assume that the conjugate distribution applies.

Procedure	Example
1. State your prior degree-of-belief.	1. $r = 4$, $T = 2000$ hr.
2. Find the equivalent failure rate and its uncertainty. See par. 2-12 for formulas.	2. $\lambda_{est,prior} = 4/2000\text{-hr} = 2/1000\text{-hr}$ coeff. of variation $= 1/\sqrt{4} = 50\%$.
3. Find the failure rate and its uncertainty, for the test results only.	3. $\lambda_{est,test} = 3/1000\text{-hr}$ coeff. of variation $= 1/\sqrt{3} = 58\%$.
4. Combine the prior and test results, per conjugate distribution theory.	4. $\lambda_{est,after} = (3 + 4)/(1000\text{ hr} + 2000\text{ hr}) = 2.3/1000\text{-hr}$. coeff. of variation $= 1/\sqrt{3+4} = 38\%$.

The prior and after-test results agree reasonably well. Before using this procedure, be sure to simulate possible test outcomes extensively. Example No. 31 shows what can happen.

Example No. 31

Same prior degree-of-belief as Example No. 30, but the test results are 3 failures in 10 hr. What is your new degree-of-belief? Assume that the conjugate distribution applies. Steps 1 and 2 are the same as in Example No. 30.

Procedure	Example
3. Find the failure rate and its uncertainty for the test results only.	3. $\lambda_{est,test} = 3/10\text{-hr} = 300/1000\text{-hr}$. coeff. of variation $= 1/\sqrt{3} = 58\%$.
4. Combine the prior and test results, per conjugate distribution theory.	4. $\lambda_{est,after} = (3+4)/(10\text{ hr} + 1000\text{ hr}) = 7/1000\text{-hr}$ coeff. of variation $= 1/\sqrt{3+4} = 38\%$.

Now, only a fool really believes that the failure rate is 7/1000-hr (about twice the prior belief). Anyone else is very worried about those test results. So, the prior degree-of-belief was most inappropriate. Perhaps $r = 0.4$ and $T = 200$ hr would have been better choices for the prior degree-of-belief.

This conjugate method of choosing prior degree-of-belief does not have the flexibility that the discrete method does, and it is easier to lead oneself astray.

easy and the descriptions are pleasantly smooth. Be extremely wary about a false sense of security. Bayesian techniques are one of the easiest ways known to mankind of making a fool of oneself. They can, and ought to be, profitably used—but not when accuracy of representation is sacrificed for mathematical tractability.

REFERENCES

1. *Reliability and Maintainability Handbook for the US Weather Bureau,* Publication 530-01-1-762, ARINC Research Corporation, Annapolis, MD, April 1967.
2. W. Grant Ireson, Ed., *Reliability Handbook,* McGraw-Hill Book Company, New York, NY, 1966.
3. M. G. Natrella, *Experimental Statistics,* National Bureau of Standards, Handbook 91, October 1966. Also, AMCP 706-110 through -114. Engineering Design Handbooks.
4. D. K. Lloyd and M. Lipow, *Reliability: Management, Methods, and Mathematics,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1962.
5. MIL-STD-105, *Sampling Procedures and Tables for Inspection by Attributes.*
6. H. T. Dodge and H. G. Romig, *Sampling Inspection Tables,* John Wiley & Sons, NY, 1944.
7. M. Sobel and J. A. Tischendorf, "Acceptance Sampling with New Life Test Objectives", *Proceedings of the Fifth National Symposium on Reliability and Quality Control,* 108-18, 1959.
8. MIL-STD-414, *Sampling Procedures and Tables for Inspection by Variables for Percent Defective.*
9. Handbook H-108, *Sampling Procedures and Tables for Reliability and Life Testing,* Office of the Assistant Secretary of Defense (Supply and Logistics), Washington, DC, 1960.
10. H. P. Goode and J. H. K. Kao, *Sampling Plans Based on the Weibull Distribution,* TR 1, Department of the Navy, Office of Naval Research, 1960.
11. B. Epstein, *Statistical Techniques in Life Testing,* Chapter III, "Testing of Hypotheses", Wayne State University, TR 3, October 1958, AD 21 145.
12. B. Epstein and M. Sobel, "Life Testing", Journal of the American Statistical Association **48,** No. 263 (September 1953).
13. I. Bazovsky, *Reliability Theory and Practice,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1961.
14. A. Wald, *Sequential Analysis,* John Wiley & Sons, Inc., New York, NY, 1947.
15. AR 702-3, *Product Assurance: Army Materiel Reliability, Availability, and Maintainability (RAM).*
16. *Sequential Analysis of Statistical Data: Applications,* Statistical Research Group, Columbia University. Report #SRG 225, Columbia University Press, NY, 1945.
17. MIL-STD-781, *Reliability Tests: Exponential Distribution.*
18. A. Hald, *Statistical Tables and Formulas,* John Wiley & Sons, Inc. New York, NY, 1952.
19. W. J. Conover, *Practical Nonparametric Statistics,* John Wiley & Sons, New York, NY, 1971.

20. Mann, Schafer, Singpurwalla, ***Methods for Statistical Analysis of Reliability & Life Tests,*** John Wiley & Sons, Inc., New York, NY, 1974.

21. D. Oksöy, "An Exact Analysis of the Method-One Maintainability Demonstration Plan in MIL-STD-471", IEEE Trans. Reliability, **R-21,** 207-211, (1972).

22. Aroian and Oksoy, "Truncated Sequential Life Tests for a 3-way Decision Procedure", IEEE Trans, Reliability, **R-22,** 54-72, (1973).

23. A. J. Duncan, ***Quality Control and Industrial Statistics,*** 3rd ed., E. D. Irwin, Inc., Homewood, Illinois, 1965.

24. R. A. Evans, "The Analysis of Accelerated Temperature-tests", ***Proc. of 1969 Annual Symposium on Reliability,*** 1969, pp. 294-302. (available from IEEE, 346 E. 47th St., New York, NY, 10017)

25. A. M. Breipohl, ***Probabilistic Systems Analysis,*** John Wiley & Sons, Inc., NY, 1970.

26. IEEE Transactions on Reliability, **R-21,** Special issue on Bayesian reliability techniques, August 1972.

27. R. G. Easterling, **"A** Personal **View** of the Bayesian Controversy in Reliability and Statistics", IEEE Trans. Rel. **R-21,** 186-194 (1972).

28. R. A. Evans, "Bayesian Analysis of Risk", Quality Progress **VIII,** 9-12 (1974). (Also published in the ***Proceedings of the Product Liability Prevention conference,*** 1973).

29. W. Yurkowsky, ***Nonelectronic Notebook,*** RADC-TR-69-458, Hughes Aircraft Company, Prepared for Rome Air Development Center, March 1970.

CHAPTER 4

TEST MANAGEMENT AND PLANNING

LIST OF SYMBOLS

α = producer risk

β = consumer risk

θ = mean failure time

θ_0 = acceptable value of 8

θ_1 = unacceptable value of 8

4-1 INTRODUCTION

The effectiveness of a reliability test program depends on the thoroughness with which the program is planned (Refs. 1, 2, and 3). Reliability tests often represent millions of dollars invested in test manpower and hardware, and require careful coordination and scheduling of hardware, test facilities, and the work of many engineering and technical personnel. Without proper planning, all of these elements may not be available when needed, tests may not be performed at the right time, or test schedules may be rushed and haphazard.

High-reliability military projects often operate on very short time schedules. Therefore, test planning must begin very early so that special test fixtures can be designed (on a calculated risk basis) and ready when the first hardware is tested.

Proposed test plans must be available to product designers so that test points can be designed into the system. Good test planning may reveal that the cost of required test procedures, test points, and test manpower is excessively high relative to the cost of the item tested. It may then be necessary to redesign the product.

Coordination between responsible groups is important in the planning of reliability testing. A frequent error made by project managers is to let the product and test equipment designers work out the details of the acceptance or quality testing of development hardware without consulting reliability and quality assurance personnel. When this occurs, reliability and quality assurance may be treated inadequately. In a well managed project, reliability and quality assurance engineers plan all reliability and quality testing with inputs provided by product design groups. This approach works well only if the reliability group begins developing the test plans concurrently with the beginning of product design.

Tests must be planned in great detail and must cover all elements of a test program in order to ensure that useful data are produced. Test planning must be complete enough to permit duplications, and omissions to be uncovered and cost trade-offs to be made.

4-2 PROGRAM PLANNING

4-2.1 MANAGEMENT ORGANIZATION FOR TESTING

The overall responsibility for planning all test programs and tests should be assigned to the project reliability group. This ensures that the requirements of design, quality assurance, and field service groups can be met in a coordinated fashion relating to reliability.

A reliability test planning committee, chaired by a member of the reliability group, should be organized and begin to function at the start of the project. This committee should meet on a regular basis to update the overall test-program plan as changes occur on the project. This committee should include members of reliability, quality assurance, product design, test equipment design, production planning, and test laboratories.

The representatives of each technical speciality make contributions to the test planning group. The design engineer specifies the items and attributes to be tested and inspected to ensure proper operation of the system ; the quality assurance representative specifies the attributes to be tested for process and quality control; the reliability engineer defines the reliability verification requirements; test equipment personnel contribute heavily to the test program planning. The committee estimates costs and assesses the overall balance between risk and cost.

The basic characteristics of development tests, qualification tests, demonstration tests, and quality-assurance tests which must be considered by test management are listed in Tables 4-1 through 4-4 (Ref. 4). Typical steps which must be followed when planning a reliability test are summarized in Table 4-5 and information categories that should be included in a test plan are described in Table 4-6.

4-2.2 SCHEDULES

Reliability test planning for all tests should begin at the very start of a project (Ref. *2)*. An error that should be avoided is to postpone planning for system demonstration tests to a point later in the project. All test planning and scheduling should begin immediately. **A** preliminary classification should be made for all proposed tests. Test plans should be recorded in spread-sheet format. These preliminary plans must be revised frequently as the program proceeds.

Reliability-schedule changes can result from many factors, including design changes, changes in production techniques or location, and changes resulting from information derived from design reviews and preliminary testing. Reliability schedules developed early enough in the program will be flexible enough to incorporate needed changes as the program progresses.

4-2.3 DOCUMENTATION

Much documentation is required to plan a full reliability-test program. The reliability-test plans should be submitted by contractors as part of their proposals to the Army and, upon approval, be included in the contract or detailed equipment specification. As development progresses, the plan is updated as required.

Table 4-6 lists the general information categories of a test plan. Further details of a reliability-test plan are:

1. Description of Demonstration Conditions.

a. Reliability requirements. The values of specified MTBF and minimum acceptable MTBF or other measures of reliability. See latest version of Ref. **7.**

b. Test purpose. (1) qualification of new or redesigned equipment; **(2)** sampling of production equipment; and **(3)** longevity test

c. Equipment identification. A detailed description of the equipment under test with notes about necessary auxiliary equipment, failure of which will not be charged to the equipment under test

d. Demonstration sites and facility requirements

TABLE 4-1

MANAGEMENT ASPECTS OF DEVELOPMENT TESTS[4]

1. Purpose of Tests
To determine physical realizability, to determine functional capabilities, to establish the basic design.

2. General Description
Development tests are usually informal exploratory tests designed to provide fundamental R&D information about a basic design. Nominal environmental levels are used unless the test is oriented specifically to check for effects at environmental extremes. Sample sizes are limited, but the general principles of good experimental and statistical design should be followed.

3. Examples of Specific Types of Tests
a. Design-Evaluation Tests
b. Fatigue Tests
c. Environmental Tests
d. Functional Tests
e. Breadboard Tests
f. Critical-Weakness Tests
g. Compatibility Tests

4. Test Scheduling
Not usually specified formally. Design-engineering group establishes schedules to meet design development objectives. Such schedules must conform to development-program milestones.

5. Test Items
Basic materials, off-the-shelf parts and assemblies, breadboard models, prototype hardware.

6. Test Documentation
Engineering test reports and analyses. Performance, failure, and maintainability information to be documented for later use in prediction, evaluation, and testing tasks.

7. Test Follow-Up Action
Determination of design feasibility or need for redesign. Implementation of test information in further design work. Approval, modification, or disapproval of design, materials, and parts.

8. Reliability/Maintainability Provisions
Proposed materials and designs to yield acceptable R&M performance are tested on limited samples. Material-fatigue tests, packaging tests, component-interaction tests, accelerated environmental tests, etc., are examples. All R&M data should be fully documented for future use in prediction, assessment, and later testing activities.

e. Participating agencies

2. Description of Test Team:

a. Organization

b. Degree of participation of contractor and procuring activity

c. Assignment of specific responsibilities

d. Qualification, quantity, and training of test-team personnel

3. Description of Demonstration Support Equipment:

a. Support equipment

b. Tools and test equipment

TABLE 4-2

MANAGEMENT ASPECTS OF QUALIFICATION TESTS[4]

1. Purpose of Tests
To demonstrate that the equipment or specified components, assemblies, and packages meet specified performance requirements under stated environmental conditions.

2. General Description
Qualification tests are formal tests conducted according to procedures specified in the development contract. Sample size is small, and thus inferential analysis is limited.

3. Examples of Specific Types of Tests
   a. Preproduction Tests
   b. Environmental Tests
   c. Functional Tests
   d. Compatibility Tests
   e. Safety-margin Tests
   f. Continuity Tests
   g. Quality Tests

4. Test Scheduling
Normally contract specified be performed before production release.

5. Test Items
Pilot-line items produced, to the extent possible, under normal production methods.

6. Test Documentation
Detailed test requirements and procedures. Test results fully documented, including analyses and conclusions concerning design qualification.

7. Test Follow-Up Action
Approval of design or implementation of recommended changes to correct deficiencies. Design approval permits production release.

8. Reliability/Maintainability Provisions
Limited reliability and maintainability assessments may be specified during design qualification tests, such as a short continuous-operation test or tests of failure diagnostic routines. Primary applications are limited, however, to quality testing of parts and processes.

---

c. Technical publications

d. Spares and consumables

e. Safety equipment

f. Calibration support requirements

4. Predemonstration-Phase Schedule:

a. Assembly of test team

b. Training

c. Preparation facilities and support material

TABLE 4-3

MANAGEMENT ASPECTS OF DEMONSTRATION TESTS[4]

1. Purpose of Tests

   To demonstrate formally that operational requirements in terms of effectiveness parameters such as reliability, maintainability, and design capability are achieved.

2. General Description

   Demonstration tests are performed on the major end items, often at the highest system level, under realistic operational and environmental conditions. Rules are specified for classifying failures, performing repairs, allowing design changes, etc. Time is an inherent test parameter. The test design is usually directed towards providing a specified s-confidence for making an appropriate decision.

3. Examples of Specific Types of Tests

   a. Reliability Demonstration
   b. Maintainability Demonstration
   c. Availability Demonstration
   d. Life Tests
   e. Longevity Tests

4. Test Scheduling

   Demonstration-test schedules are normally contract-specified. They generally occur before full-scale production but after initial production, when test samples are available.

5. Test Items

   Production hardware at major end-item level.

6. Test Documentation

   Contract-specified procedures or clause requiring contractor to submit complete test plan. Test results fully documented, including analyses and conclusions concerning the meeting of contract goals.

7. Test Follow-Up Actions

   Acceptance or rejection of equipment with respect to reliability, maintainability, and effectiveness goals. Failure to pass demonstration tests will require appropriate design and assurance efforts on the part of the contractor.

8. Reliability/Maintainability Provisions

   Demonstration tests are specifically designed to test for reliability, maintainability, and associated parameters at the equipment level. Demonstration tests may be continued throughout the production cycle on samples of equipment.

---

5. Description of Formal Demonstration Test:

a. Performance parameters to be measured

b. Performance limits for defining failure

c. Preventive-maintenance measures to be performed

d. Definition of minor failures (not to be included in MTBF analysis)

e. Test equipment to be used

f. Test severity levels (see par. **4-3.2)**

g. Monitoring test equipment

h. Corrective-maintenance procedures

TABLE 4 4

MANAGEMENT ASPECTS OF QUALITY-ASSURANCE TESTS[4]

1. Purpose of Tests
Quality-assurance tests are performed on samples of incoming or outgoing products to assure that materials, parts, processes, and final product meet the established performance and quality levels.

2. General Description
Quality-assurance tests, performed during the production phase, include two basic types: (1) acceptance tests on samples of items, to accept or reject a lot; and (2) quality-control tests on processes and machines, to ensure that final product will be satisfactory. The tests are usually designed on a statistical basis to meet specified risk levels.

3. Examples of Specific Types of Tests

a. Percent-defective test	e. Incoming-inspection tests
b. Parts-screening tests	f. Storage tests
c. Production-control tests	g. Machine-wear tests
d. Part-lot acceptance tests	h. Continuous-sampling tests

4. Test Scheduling
Quality-assurance tests are scheduled throughout the production phase, on either a lot-by-lot basis or on a continuous basis, depending on the circumstances. Scheduling of tests can depend on past performance of the Contractor.

5. Test Items
Incoming material, machines that process the material, and production end items at all levels.

6. Test Follow-Up Action
Acceptance or rejection of processes or production lot. Rework of rejected lots may be provided for. Many plans tighten risk levels of poor producers, or relax levels if good quality is maintained.

7. Reliability/Maintainability Provisions
A reliability acceptance test on go/no-go items is a normal quality-assurance test. Time tests for testing mean life may be scheduled periodically but may not be as extensive as the initial demonstration tests. Maintainability usually is tested only indirectly.

i. Data-analysis and calculation measures

j. Time units of measurement

k. Type and schedule of report and log forms

6. Retest Phase. Provisional retest schedule.

4-2.4 TEST PROCEDURES

Well prepared test procedures are extremely important for an effective test program. Detailed, formal, and controlled test procedures are required in high reliability programs and complex weapon system programs.

Test procedures describe and control (1) test equipment calibration, (2) test equipment proofing, and (3) test programming. An elaboration on the elements follows:

1. Calibration. Test equipment should be calibrated against standards traceable to the National Bureau of Standards. Calibration operations should be performed at the interfaces between the test equipment and the hardware (at the test leads). All test equip-

TABLE 4-5

STEPS IN OVERALL TEST PLANNING[4]

1. Determine test requirements and objectives.
2. Review existing data to determine if any existing requirements can be met, without tests.
3. Review a preliminary list of planned tests to determine whether economies can be realized by combining individual test requirements.
4. Determine the necessary tests.
5. Allocate time, funds, and effort to perform these tests.
6. Develop test specifications at an appropriate level, or make reference to applicable sections of the system specification to provide direction for later development of test specifications.
7. Assign responsibility for test conduct, monitoring, analysis, and integration.
8. Develop review and approval policies for test-reporting procedures and forms.
9. Develop procedures for maintaining test-status information throughout the entire program.

ment should be calibrated, including measuring equipment and environmental test chambers. At the start of a test, calibrations should be checked over the range of values expected.

2. Proofing. The test procedures must provide techniques for demonstrating that test equipment will function properly when coupled with test hardware. These procedures are known as proofing. They can be used to uncover unanticipated problems such as ground loops and variations in input conditions with variations in loading. Proofing is very important the first time a test equipment design is used with a specific set of hardware.

3. Programming. Test programming includes all the minute operations required of the test personnel and equipment. Detailed data sheets that define all the required input and output data and their units are required. The data sheets should include spaces for recording nontest information—such as laboratory environmental conditions, date, hardware configuration, test operator and inspector identification, and other administrative data. Acceptance-test accept/reject limits also should be included on the data sheets.

A system of controls and check-off procedures should be established so that the test plan and changes are reviewed and approved by all interested parties.

4-3 TEST CRITERIA

4-3.1 SELECTION OF ATTRIBUTES

The selection of attributes for testing depends on many factors, such as (Ref. 2):

1. The need to demonstrate that a system is functional
2. The need to demonstrate reliability
3. The cost of testing
4. The test time required
5. Equipment and personnel available for the tests
6. Army requirements
7. Requirements for repair part interchangeability
8. Desire to provide optimum process and quality control and to assure repeatability of the production processes
9. Reliability requirements

TABLE 4-6

INFORMATION CATEGORIES FOR A DEVELOPMENTAL TESTING PLAN[4]

Information Category	Description
Quantity	Number of test specimens to be built or purchased
Date	Dates of delivery of test specimens and test equipments; dates on which testing is to commence and conclude
Test Duration	Expected length of time testing is to continue on an item
Test Type	Examples: test to failure, nondestructive test, life test, etc.
Environments	Stresses to be imposed and cycling rates; parameters to be monitored; applicable specifications (e.g., military or detailed equipment specifications)
Test Procedures	Applicable specifications (e.g., military); frequency and type of monitoring required; definitions of failure or satisfactory operation; repair actions to be allowed
Test Location	Place(s) where testing is to be performed
Cost of Testing	Costs of test specimens and special test equipment needed; total cost of test
Reporting	Frequency of interim reports; types of analyses to be prepared; data forms to be employed and their distribution; allowable delay between test completion and issuance of final test report; distribution of test reports
Responsibilities	Specific personnel (or group) obligations for preparation and design of test plans and procedures, procurement of test specimens and test equipment, operation of tests, analysis of test data and results, and preparation of interim and final reports
Reliability Requirement	Statement of reliability level to which equipment will be tested
Maintainability Requirement	Statement of maintainability level to which equipment will be tested

10. The procurement cost and the cost of replacement.

It is not possible to test all attributes of a component or a system. Therefore, only a subset of the attributes describing a system can be tested. The process of selecting those attributes to be tested requires the exercise of a great deal of judgment. The committee approach can be very useful.

The process of developing a list of attributes for test is simplified if a standard classification procedure is used. A system of classification of defects and classification of characteristics was developed for use by the Army and Navy (Ref. 5). In this system, each attribute is classified critical, major, or minor, in accordance with its effect on coordination, life, interchangeability, function, and safety. With such a classification system in operation, attribute classification is standardized from project to project, from item to item, and from test program to test program, within a project. The use of a standard classification system simplifies failure diagnosis, corrective action, inspection, and design-change control, and provides baseline definitions for the reliability incentive in incentive contracts. The most direct benefit of test planning is that subjective selection of attributes is replaced by objective application of an agreed-upon set of ground rules.

4-3.2 TEST CRITERIA FOR RELIABILITY DEMONSTRATION PER MIL-STD-781* (REF. 6)

4-3.2.1 Test Levels

MIL-STD-781 has 10 different test levels that specify conditions of temperature and temperature cycling, input voltage cycling, on-off cycling, and vibration. Table 4-7 summarizes the test levels. These levels should be considered minimum requirements, and appropriate modifications should be made to meet more stringent conditions. The following considerations should apply:

1. The test level should be severe enough to equal the anticipated operational stress.

2. The test level must be severe enough to uncover defective parts or workmanship.

3. The test level should approach or equal design extremes but should not exceed basic design specifications to such an extent that non-relevant failure modes become important.

4-3.2.2 Test Criteria

MIL-STD-781 provides five types of tests: (1) standard sequential tests; (2) short-run, high-risk sequential tests, (3) fixed-length tests; (4) longevity tests; and (5) all equipment screening test. There are 29 individual test plans. Their risk characteristics are summarized in Table 4-8.

Generally, the standard sequential plans (I through VI) offer acceptable risk levels at minimum test time for development tests. The high-risk sequential tests (plans VII, VIII, IX) should be used only if test resources are very limited and the high risks of incorrect decision are acceptable. The fixed-length tests (X through XXV) are generally used for production sampling since the accept test time is fixed, leading to easier scheduling, and since, generally, more test equipment is available for production testing than for developmental testing. Plan XXVI (development) and XXVII (production) specify a maximum test time of 500 hr on each sample equipment. Decisions can be made after a period equal to 3 times the specified MTBF (if such a period is less than 500 hr).

The longevity test is used for testing the total operational life of the equipment. At least two equipments must be tested for a time equal to the specified longevity. Time accumulated on the demonstration test may be applied to the longevity test. If there is no longevity requirement, each equipment is tested for 2000 hr. No accept criteria are given, but all failures and patterns are analyzed to determine if the longevity goal is satisfied.

4-3.2.3 Test Performance and Evaluation

The following items are important:

1. Sample Size. Generally at least 3 equipments are tested, but the actual number depends on the purpose of the test and the lot size. See Ref. 6.

2. Evaluation Criteria. MIL-STD-781 presents the complete accept-reject criteria for each of its 29 test plans (i.e., the ones shown in Table 4-8).

3. Test Procedure. MIL-STD-781 pre-

*All references to MIL-STD-781 are to version B, Change 1, July 1969 revision.

TABLE 4-7

SUMMARY OF TEST LEVELS
(Adapted from Ref. 6)

Test Level	Temperature	Temperature Cycling See Note	Vibration See Note	Equipment On-Off Cycling See Note
A	25±5	None	1	2
A-1	25+5	None	None	None
B	40±5	None	1	2
C	50+5 /–0	None	1	2
D	65±5	None	1	2
E	–54 to 55	5	1	6
F	–54 to 71	5	1	6
G	–54 to 95	5	1	6
H	–65 to 71	5	1	6
I	–54 to 125	5	1	6
All	See Note 3 for Input Voltage, and Note 4 for Input Voltage Cycling.			

Notes

1. 2.2G ± 10% peak acceleration value at any nonresonant frequency between 20 and 60 Hz measured at the mounting points on the equipment. The duration of vibration shall be at least 10 min during each hour of equipment operating time.

2. Turn on and let temperature stabilize, hold for 3 hr, then turn off and let temperature stabilize. This cycle shall continue throughout the test.

3. Nominal specified voltage plus 5%, minus 2%.

4. When so directed by the procuring activity, voltage cycling shall be accomplished as follows: The input voltage shall be maintained at 110% nominal for one-third of the equipment "on" cycle, at the nominal value for the second one-third of the equipment "on" cycle, and at 90% for the final one-third of the equipment "on" cycle. This cycling procedure is to be repeated continuously throughout the reliability test.

5. Temperature cycling shall be: time to stabilize at low temperature followed by time to stabilize at the high temperature, plus 2 hr.

6. Equipment off during cooling cycle and on during heating cycle.

sents procedural guidelines for selecting and installing equipment, initiating tests, heating and cooling cycles, repeated testing, determining compliance, failure actions and failure categories, failure analysis and information, verifying repair, preventive maintenance and corrective action, restoration of failed equipment, test records and reports, and final reports.

4-4 TYPICAL ARMY SCHEDULE

Fig. 4-1 shows the test support for material acquisition.

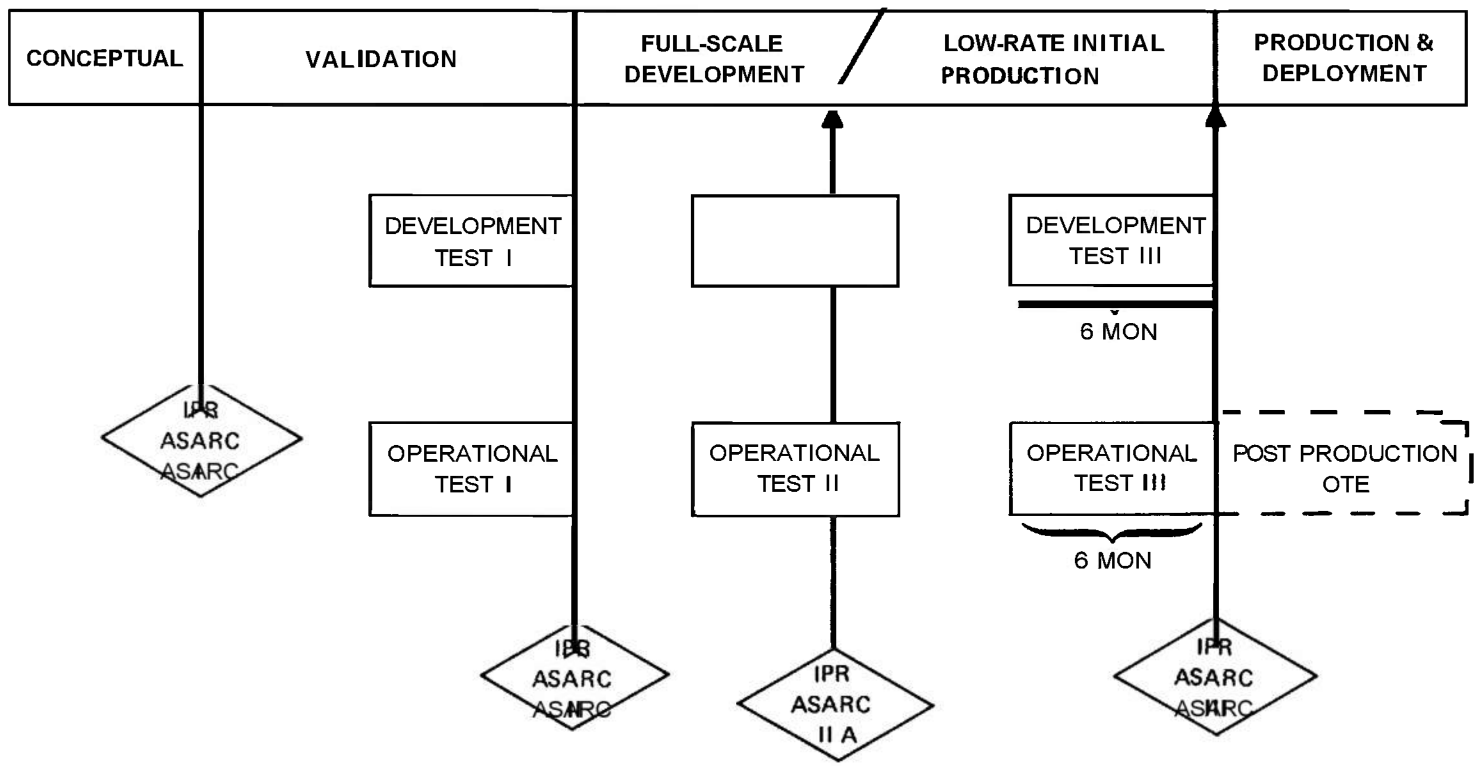
TABLE 4-8

SUMMARY OF RISK AND TIME CHARACTERISTICS FOR INDIVIDUAL TEST PLANS FOR CONSTANT FAILURE-RATE EQUIPMENT (ADAPTED FROM Ref. 6)

θ_0 = true mean life at producer risk point
θ_1 = true mean life at consumer risk point

Test Plan Number	Nominal Producer Risk α, %	Nominal Consumer Risk β, %	Discrimination Ratio θ_0/θ_1	Approximate Maximum Duration (units of θ_0)	Approximate Expected Test Time to Accept if $\theta=\theta_0$ (units of θ_0)
	actual given in ( )				
			Standard Tests		
I	10 (11.5)	10 (12.5)	1.5	33.0	17.3
II	20 (22.7)	20 (23.2)	1.5	14.6	7.6
III	10 (12.8)	10 (12.8)	2.0	10.3	5.1
IV	20 (22.3)	20 (22.5)	2.0	4.9	2.4
IVa	20 (18.2)	20 (19.2)	3.0	1.5	1.1
V	10 (11.1)	10 (10.9)	3.0	3.5	2.0
VI	10 (12.4)	10 (13.0)	5.0	1.3	0.64
			Short-Run, High-Risk Sequential Tests		
VII	30 (31.9)	30 (32.8)	1.5	4.5	3.4
VIII	30 (29.3)	30 (29.9)	2.0	2.3	1.3
IX	35 (36.3)	40 (39.7)	1.25	8.3	5.0
			Fixed-Length Tests		**Acceptance Number**
X	10	10	1.25	100	111
XI	10	20	1.25	72	82
XII	20	20	1.25	44	49
XIII	30	30	1.25	15	16
XIV	10	10	1.5	30	36
XV	10	20	1.5	20	25
XVI	20	20	1.5	14	17
XVII	30	30	1.5	5.3	6
XVIII	10	10	2.0	9.4	13
XIX	10	20	2.0	6.2	9
XX	20	20	2.0	3.9	5
XXI	20	30	2.0	1.8	2
XXII	10	10	3.0	3.1	5
XXIII	10	20	3.0	1.8	3
XXIV	20	20	3.0	1.5	2
XXV	30	30	3.0	0.37	0
XXVI	N/A	N/A	N/A	3.0	*
XXVII	N/A	N/A	N/A	3.0	*
XXVIII	longevity test plan*				
XXIX	all equipment screening test*				

*See Ref. 6 for details.



ASARC = ARMY SYSTEMS ACQUISITION REVIEW COUNCIL
IPR = IN PROCESS REVIEW

Figure 4-1. Test Support of Materiel Acquisition[8]

The Department of Army (DA) and US Army Materiel Command (AMC) continually update their practices and schedules in order that learning from past experience and present thinking can improve the materiel acquisition process in the future. It is not feasible to list current directives and thinking since they change from time to time. Therefore, one should contact the appropriate Directorates of AMC for current information.

REFERENCES

1. *Testing, Practical Reliability,* Vol. III, NASA CR-1*128,* Research Triangle Institute, Research Triangle Park, NC, August 1968.

2. W. Grant Ireson, Ed., *Relubility Handbook,* McGraw-Hill Book Company, New York, NY, 1966.

3. *MIL-STD-785, Requirements for Reliability Program (for Systems and Equipments), 1965.*

4. *Reliability and Maintainability Handbook for the US Weather Bureau,* Publication *530-01-1-762,* ARINC Research Corporation, Annapolis, MD, April 1967.

5. Bureau of Ordnance Standard 78.

6. MIL-STD-781, *Reliability Tests, Exponential Distribution,* 1969.

7. AR 702-3, *Product Assurance: Army Materiel Reliability Availability and Maintainability (RAM).*

8. Henry Meodozeniec "The Materiel Acquisition Process for DA" presented at Joint AMC/TRADOC RAM Seminar, May 1974 at AMC Headquarters.

CHAPTER 5

INTEGRATED RELIABILITY DATA SYSTEM

5-1 INTRODUCTION

An integrated reliability data system can be used to provide project managers and engineers with the data that they need to determine the reliability achieved by the system and its component parts. If provided in a timely manner, this information can be used for effective planning, review, and control of actions related to system reliability.

The data system should be established for collecting, analyzing, and recording data on all failures that occur during system development and operation. The system must provide data that can be used to estimate reliability and from which needed corrective action can be determined. Computer programs that permit the printing of reliability status reports should be developed or acquired as part of the integrated reliability data system.

The reliability data system will be useful to the designer in providing a complete failure history of the system and its constituent parts in some easily interpreted form. This history should include an indication of the specific mode of failure, the cause of each failure, and a record of the effectiveness of each corrective action.

The reliability data system also must serve management. Management must be provided with summary reports describing the current reliability status of the system. Suppliers of components must be evaluated continuously to ensure that their products have adequate reliability. Therefore, the data system must provide a current record of the failure history attributable to each vendor.

Procedures for data accumulation and reduction must be developed and standardized. These standard procedures must provide for the collection of data – such as identifying information, environmental conditions, operating hours or cycles, and the description of hardware failures on each test performed. The system also should be structured to make use of data recorded on failures occurring at times other than the reliability tests.

The integrated data system can be used to handle, process, and integrate all data obtained from testing, inspection, and failure-trouble reporting. These data can be used for reliability analysis and reporting, assessment of equipment readiness, and a variety of other purposes.

A computer data bank of accumulated and summarized reliability data must be maintained and updated periodically. These data can be processed to produce reliability parameters for components, equipments, and subsystems. These reports can be structured to present a listing of troublesome items causing the most serious reliability problems. These lists then can be distributed to cognizant Army and contractor engineers and managers.

The reliability data system for a weapon system developed by the Army should be usable by both contractor and Army personnel. During research, development, and engineering design, the data system provides the information required for reliability design decisions. Later, when the system becomes operational, the data system can be used by the Army to collect field reliability data (if so desired) which will provide the basis for system modifications and changes in maintenance and supply concepts. The DA operates several data banks; see Part Two, App. B for a listing of some of the data bank/retrieval systems. The policy in regard to data banks changes occasionally. Check the latest directives in this regard.

5-2 STRUCTURE OF A RELIABILITY DATA SYSTEM

A reliability data system consists of a data bank, a set of computer programs, and computer hardware. The data bank is a systematic set of data describing the reliability characteristics of selected component parts and subsystems as well as the system as a whole. The information in the data bank can be made available in printed form in a variety of formats, using report generator programs. The information in the data bank also can be manipulated to produce reports required by systems management and engineering personnel. The use of a single unified data base simplifies the problems of file searching, report generation, and adding new data.

In this paragraph, the following areas will be discussed: (1) organizing and addressing data, and data format, (2) programs for data bank establishment and updating, (3) extraction routines, and (4) data bank operation.

5-2.1 ORGANIZING AND ADDRESSING DATA

The information in the data bank must be organized in a systematic manner, so that it is uniquely addressable and readily accessible. The information retrieval programs should permit standard reports to be generated and specific questions to be answered on query.

Once a suitable organizational structure for data has been determined, a corresponding address structure must also be devised so that an address is available to designate uniquely any category in the organizational hierarchy.

The data elements to be stored in the system must be carefully defined and structured prior to establishing the files. A typical data element structure is given in Table 5-1 (Ref. 3). This is suitable only for testing up to, but not including, ordinary field use. Standard DA procedures are to be used for field failure reporting (Refs. 1, 2).

The classification system should permit reliability data to be organized by system, subsystem, assembly, subassembly, and lowest replaceable unit. A numerical coding scheme should be developed which permits the system hierarchical structure to be described.

The coding system also must permit failure modes and test environments to be described. All environmental factors to be applied during the reliability and environmental tests, as well as those factors expected to be encountered in the field, must be included in the system. Fig. 5-1 (Ref. 4) shows a typical computer printout summarizing a typical set of environmental factors. As the project progresses, this list can be expanded.

The data formats most suitable for computer manipulation may not coincide with formats that are easy for the test engineers to use: therefore, the formats for data entry may be somewhat different than those used by test engineers. The way in which data reformatting is accomplished depends on the system. For example, one approach is to reorganize the data when entering it onto data-entry coding forms. Or, the data can be entered directly from test forms (after error checking) to be reformatted by the computer.

TABLE 5-1

DATA ELEMENT DEFINITIONS AND IDENTIFICATION OF SYSTEM SOURCES OF DATA APPLICABLE TO EACH DATA ELEMENT DESCRIPTOR[3]

Data Elements	Definition or Explanation
ITEM DESCRIPTORS	
(1) Item Identification	
FSN/Bureau Plan and Piece Number, Drawing Number	Federal stock number, bureau piece number, or drawing number of system equipment
CID/APL/AN Number	Component identification number, Allowance-Parts-Lists number, Army-Navy Number of equipment in which replacement part was used
System/Equipment Name	Noun name identification of system/equipment at the highest assembly level
System/Equipment Part Number or Identification Code	Federal stock number (FSN) at the highest assembly level
System/Equipment Serial Number	Manufacturer's serial number assigned to the system/equipment
Vehicle Serial Number	Serial number of missile, aircraft or other vehicle in which failed part was located
Assembly Name	Noun name identification of the assembly in which the failed part is located
Assembly Part Number or Identification Code	Federal stock number (FSN) of the assembly
Assembly Serial Number	Manufacturer's serial number of the assembly containing failed part
Subassembly Name	Noun name identification of the subassembly where the failed part is located
Subassembly Part Number or Identification Code	Federal stock number (FSN) of the subassembly containing failed part
Subassembly Serial Number	Manufacturer's serial number of the subassembly containing the failed part
Subassembly Symbol/Designation	Manufacturer's drawing reference, circuit, symbol, or other identification of the subassembly containing failed part
Failed Part/Item Number	Federal stock number (FSN) of the failed part
Failed Part/Item Name	Noun name identification of the failed part
Failed Part/Item Serial Number	Manufacturer's serial number of the failed part or item

TABLE 5-1 (Cont'd)

DATA ELEMENT DEFINITIONS AND IDENTIFICATION OF SYSTEM SOURCES OF DATA APPLICABLE TO EACH DATA ELEMENT DESCRIPTOR[3]

	Data Elements	Definition
	ITEM DESCRIPTORS (Cont'd)	
	Failed Part/Item Symbol Designation or Code	Manufacturer's drawing reference, circuit, symbol or other identification of failed part or item
	Federal Stock Number (Removed Item)	Federal stock number (FSN) of failed parts of items removed from equipment
	Part Number (Installed Item)	Part number or federal stock number of replacement part or item
	Serial Number (Installed Item)	Manufacturer's serial number of replacement part or item
(2)	Hardware Location and Source Identification:	
	Location (Geographic)	Location of the equipment that is the source of the data
		Geographic location of equipment when part failed
	Location (Physical)	Location of failed part in the equipment, or name of assembly if more than one of the same part is used
	Installed in A/C Arresting Gear, Catapult, or Support Equipment	Model description and serial number of equipment in the categories where failed part was located
	Equipment Contractor	Name of contractor or manufacturer of equipment
	System/Equipment Manufacturer	Noun name identification of prime manufacturer
	Assembly Manufacturer	Noun name identification of manufacturer of assembly containing failed part
	Subassembly Manufacturer	Noun name identification of manufacturer of subassembly containing failed part
	Failed Part/Item Manufacturer	Noun name identification of manufacturer of the failed part or item
	Manufacturer Name or Code (Component/Assembly Replacements)	Noun name identification of manufacturer of replacement assembly
	Manufacturer Name or Code (installed Item)	Noun name identification of manufacturer of replacement part of item

TABLE 5-1 (Cont'd)

DATA ELEMENT DEFINITIONS AND IDENTIFICATION OF SYSTEM SOURCES OF DATA APPLICABLE TO EACH DATA ELEMENT DESCRIPTOR[3]

Data Elements	Definition or Explanation
ITEM DESCRIPTORS (Cont'd)	
Contract Number	Number identification of contract under which the system/equipment containing failed part was procured
Production Status	Development, preproduction, or operational at the time failure occurred
RELIABILITY DESCRIPTORS	
(1) Time and Number Data Elements:	
Date Form Submitted or Date of Report	Calendar date form is submitted or calendar date report is completed
Date of Failure	Calendar date failure occurred or malfunction first observed
Time of Failure	Clock time failure occurred or was first observed
Date of Last Failure	Calendar data of last failure of any kind on the equipment
Total System/Equipment Operating Time	Total clock hours of operating time logged on the equipment when the failure occurred
Operating Hours on Failed Part	Total clock hours of operating time accumulated on the failed part
Time Meter Readings (Log Book Time/Malfunctional Equipment)	Clock hours of operating time on the equipment – from meters or log book – when failure occurred
Operating Hours Since Last Component Failure	Total operating hours on failed equipment since the last part failure of any kind
Miles	Mileage from odometers mounted on the equipment where failed part is located
Rounds	Total number of rounds fired
Starts	Total number of hot starts for jets or turbine engines
Time Since New (Vintage) of Equipment (Year of Operating Status)	Calendar years and months since the equipment was installed in a new condition for operational use
Equipment Downtime	The total time during which the equipment was not in acceptable operating condition

TABLE 5-1 (Cont'd)

DATA ELEMENT DEFINITIONS AND IDENTIFICATION OF SYSTEM SOURCES OF DATA APPLICABLE TO EACH DATA ELEMENT DESCRIPTOR[3]

Data Elements	Definition or Explanation
RELIABILITY DESCRIPTORS (Cont'd)	
Total Systems (Number)	Total numbers of equipments in operation at the reporting activity
System Mean Time	Total measured operating time divided by the total number of failures
Total Number of Failures	The sum of all failures involved in an equipment malfunction
Number of Failures (Each Mode)	Total number of failures in each failure mode for each equipment malfunction
Number of Failures (Each Part)	Total number of parts as related to total number consumed in making the repairs
Failed Material (Quantity)	Total number of parts replaced during each equipment malfunction
Estimated Percent of Total Failures Reported	Estimated percent of all failures reported during a given report period
Failure Rate	At any point in the life of material, the incremental change in the number failures (change in the measure of life)
Part/Component Population	Total number of parts or components in a given universe under study
Date and Duration of Test	Calendar date of test where failure occurred and duration in hours to the time of failure
(2) Circumstantial Data Elements:	
Identification of Test or Activity in Progress	Conditions of test, type of test, test data, and duration
Status of Equipment	Circle an arrow to indicate status of equipment prior to incorporating specified technical directives
Intended Use	Intended end use environment by installation environment
Environment	Environment when failure occurred
Special Environmental Conditions	Special environmental conditions when failure occurred
Failure Reporting System	Controlled or uncontrolled system, method of reporting, personnel reporting, definition of failure, and estimated percent of total failures reported

TABLE 5-1 (Cont'd)

DATA ELEMENT DEFINITIONS AND IDENTIFICATION OF SYSTEM SOURCES OF DATA APPLICABLE TO EACH DATA ELEMENT DESCRIPTOR[3]

Data Elements	Definition or Explanation
RELIABILITY DESCRIPTORS (Cont'd)	
Type of Report	Check-off list of six classes of reports; approximate block is checked to indicate type of report
Report Priority	The assignment of priority classifications such as "Urgent" and "Flight Safety" to the report
Equipment Status After Failure	Equipment performance after failure occurred
Type of Failure (Critical/Major/Minor)	The one code, out of three, best describing the type of failure
Primary or Secondary Failure	To indicate a prime failure or a failure caused by the failure of another part
Operational Condition	One of three classifications describing effect of failure on equipment operation
Discovered (Code/Time/Situation)	A single-letter code which identifies when malfunction of the equipment or component was discovered
Symptoms (Description of Failure and Discovery/Symptom Code)	Description of any obvious reason for failure or abnormal manifestations in operation at the time of malfunction
Malfunction Description	Describes the trouble in the system, component identified in the work-unit code block
Percent of Rating (Voltage/Power, etc.)	Percent of rated load for the part application under operating-stress conditions
Description/Remarks (Additional Information)	Any additional descriptions, remarks, or suggestions related to the malfunction
Part Condition (Failed Part)	A three-digit number code describing residual condition of failed part by code system
Malfunction/Failure Cause	Cause of malfunction or failure
Failure Code	Enter code from "part condition" which best describes residual condition; may be physically observed or apparent during test or operation
How Malfunctioned	A three-digit number used to provide a description of the trouble on or in the equipment or the component listed in the FSN block

TABLE 5-1 (Cont'd)

DATA ELEMENT DEFINITIONS AND IDENTIFICATION OF SYSTEM SOURCES OF DATA APPLICABLE TO EACH DATA ELEMENT DESCRIPTOR[3]

Data Elements	Definition or Explanation
COST ACCOUNTING DESCRIPTORS:	
Contract Number	Number identification of contract under which the system/equipment containing failed part was procured
Total Systems (Number)	Total numbers of equipments in operation at the reporting activity
Total Number of Failures	The sum of all failures involved in an equipment malfunction
Number of Failures (Each Mode)	Total number of failures in each failure mode for each equipment malfunction
Number of Failures (Each Part)	Total number of parts as related to total number consumed in making the repairs
Failed Material (Quantity)	Total number of parts replaced during each equipment malfunction
Unit Cost	Unit price of parts or material used in the maintenance action, except pre-expended bin material
Estimated Percent of Total Failures Reported	Estimated percent of all failures reported during a given report period
Failure Rate	At any point in the life of material, the incremental change in the number of failures (change in the measure of life)
Part/Component Population	Total number of parts or components in a given universe under study
Maintenance – Total Man-Hours	Total man-hours required during a maintenance action
Maintenance Time – Diagnosis	Total number of man-hours required to identify cause of malfunction and determine corrective action
Maintenance Time – Active	The sum of total maintenance man-hours to diagnose failure and total maintenance man-hours for active repair
Logistics and Administration Time	Total number of man-hours repair is delayed **solely** in waiting for a replacement part and that portion of downtime not included under active repair time
Required Material (Quantity)	The number **of** units of parts or material used to accomplish a specific maintenance action
Quantity (Number of Items Received or Returned)	The number of units of parts or material used to accomplish a specific maintenance action

TABLE 5-1 (Cont'd)

DATA ELEMENT DEFINITIONS AND IDENTIFICATION OF SYSTEM SOURCES OF DATA APPLICABLE TO EACH DATA ELEMENT DESCRIPTOR[3]

Data Elements	Definition or Explanation
COST ACCOUNTING DESCRIPTORS: (Cont'd)	
Items Processed (Number)	The number of times collective action was taken against the item described in the work unit code block
Maintenance Control Number/ Job Control Number/Report Serial Number/Ship Account Number	Four-digit number assigned by the maintenance-data-control center
Disposition of Removed Item	One of 8 codes checked to indicate disposition of removed item
Repairman and Specialty/ Rate (title)	Name of personnel making repairs or adjustments to failed equipment and title or technical rating
Person Reporting — Rate (title)	Name signature of personnel recording data on report

CAPACITORS, FIXED, TANTALUM, SOLID, HERMETICALLY SEALED, CHASSIS MOUNT

REFERENCED MIL SPECS/STDS: (A) MIL-C-39658 (B) MIL-STD-2025 (C) (D) (E) (F) (G)

ANALYSIS NO.

DATE: DAY | MO. | YR

FIELD NO 1A 1B 2A 2B 3A 4A 4B 5A 6A 7A 7B 8A 8B 8C 9A 9B 10A 10B 10C 10D 10E 10F 10G 10H 10I 11A 12A 13A 14A 15A 16A 17A 18

CODING FOR FIELD 9B — TYPE(S) OF RADIATION: NUCLEAR: A - FAST, B - SLOW, C - THERMAL; ELECTROMAG: D - GAMMA, E - X-RAY; CHG PART: F - PROTON, G - ELECT, H - MISC.

OR LINE ENTRY	TEMPERATURE OPER/STOR LO °C	OPER/STOR HI °C	CYCLE/SHOCK LO °C	CYCLE/SHOCK HI °C	ALTITUDE FT X 10³	VIBRATION MAX FREQ CPS	G RMS	G²/CPS	DB (ACOUSTIC)	LIN ACCEL G MAX INTENSITY	DURATION SEC	SHOCK G MAX INTENSITY	DURATION MSEC	TOTAL NO BLOWS	TOTAL ABS	RADIATION TYPE - SEE REMARKS	LIFE TEST OPER TIME PER PC HR/CYCL	SAMPLE SIZE NO PC	FAIL PER 10⁶ PART HOURS	°C EXTERNAL CONTL TEMP	% PRESSURE	% CURRENT	% FREQ	% POWER	% VOLTS	SALT SPRAY	MOISTURE RESIST	HUMIDITY	SAND-DUST	IMMERSION/LEAK	RADIO INTERFER	EXPLOSION/BURST	TOTAL NO PARTS TESTED	REMARKS REFER TO FIELD NO.	REMARKS
A			54C	74C	35	2,000	30										500H	23		+85					100			A					60	8C	RANDOM FALL 6 FT ON CONCRETE
B			18C	70C													2,000H	57							100								57	101	50% + 75% RATED VOLTAGE APPLIED
C	55	125																									A						15	12A	D-C LEAKAGE LESS 150% INITIAL
D	55	125																							116								15		
E	55	125																									A						24	12A	PARA 5.3.1 OF TEST PROCED #9
F	55	125																									A						24	12A	PARA 5.3.1 OF TEST PROCED #9
G			54C	37C																													17	2A2B	-8CY HI TEMP INCREMENTS OF 15C
H	55	125			80					50	60	5	011	48			2,000H	12		+125							Z						15	12A	MIL-C-3965 F3-DC LEAKAGE CAPDF
I						2,000	60																										3	4B	VIB IN TWO AXES OF 35G AND 50G
J	55	125																									A						15		
K	55					2,000	15										100H	10		+65													25	1A	MIL-C-3965B
I	65		125C	85C																								A					10	1A	24 HOURS CYCLED 95% HUMIDITY

Figure 5-7. Sample Printout Tabulating Environmental Exposure[4]

(Reprinted from Reliability Handbook *with permission* of *McGraw-Hill Book Company, Inc.)*

A typical coding scheme is one developed for the COFEC reliability data system (Ref. 4). Failure causes and corrective actions are coded. using six digit positions. In this system reporting code, the first two digits are used to identify the failure mode, the second two digits specify the cause of failure, and the last two digits indicate the corrective action. Each category can occur in 99 different ways. A master code list is used which defines each failure mode, cause, and corrective action. This list can be expanded as the project progresses, so that the terms need not be defined in advance. A typical master code list is shown in Fig. 5-2.

5-2.2 COMPUTER PROGRAMS FOR DATA BANK ESTABLISHMENT AND UPDATING

Computer programs must be written or acquired which can be used to establish and update the data bank. The program details depend. of course, on the computer on which the system is implemented. Sort and merge routines must be available for sorting new data into the desired sequence and merging it with the existing file to produce a new file. Programs written in a report generator language such as RPG permit the structure of the data base to be altered and information to be added or deleted as desired. Generally, one ought to use existing programs as much as feasible; they ought to be programs already running on the computer available to you.

5-2.3 EXTRACTION ROUTINES AND PROGRAMS

Programs must be developed or acquired which output prescribed formats using selected parameters obtained from the data bank. Outputs to paper or files can be prepared by the computer through the use of appropriate extraction routines and the information stored in the data bank. The information in the data bank must be outputted in various formats convenient for use by both contractor and Army program managers.

A separate fixed-format routine may be written for each report format required. When the number of different formats is small, this procedure is economical. As a system design progresses, however, reporting requirements change. The varying reporting requirements and the variations in the number of groups using different output formats can greatly expand the total number of formats needed. Under these circumstances, the expense of programming a new extraction routine for each new output format can become excessive.

To make a separate extraction routine unnecessary for generating each output format, a variable-format extraction program can be employed. Such a program compensates for the greater programming expense involved with the ability to replace a number of fixed-format extraction routines: With a variable-format extraction program, personnel not trained in computer programming can write requests for a variety of report formats in some form of a user language. A user language can be developed specifically for the system in question or a commercially available system, such as RPG. can be used. Variable-format extraction programs can use standard output formats. However, the specific type of information listed in any output is established by the user, filling out a standard request form. Outputs thus may be tailored easily to the exact needs of the user.

5-2.4 PROGRAMMING EXTRACTION ROUTINES

Two general types of report generating programs can be developed: (1) those using fixed-format extraction routines, and (2) the automatic variable-format extraction and accumulation program. Each of these two programs will be discussed.

```
MASTER CODE LIST  for  Transducers

               prefix  T

01  Physical Discrepancy
02  Static Error Out of Tolerance, Receiving & Inspection Test
03  Erratic Output


11  Resistance Measurement Out of Tolerance


16  Leaking

.
```

```
02          Static error out of tolerance, receiving & inspection test
02 01       Drive link bali out of wiper arm ball socket
02 01 [01]  Vendor XYZ Mfg. Co. hss enlarged socket and h2s added potting
            compound for strength


02 03       Leak atBourdon tube braze joint
02 03 [01]  Vendor XYZ Mfg. Co. operators are trained at -------
            Brazing School effective Jan.12,63. serial no.xxx
02 03 [02]  Vendor XYZ Mfg. Co. now using improved leak test
            effective Mar.3,63.

.
.
02 03 [01]  Vendor ABCD Mfg. Co. has revised brazing process
            effecttve serial no.xxx
02 03 [02]  Vendor ABCD Mfg. Co.  has redesigned brazed section
            effective serial no. 152
```

Figure 5-2. Typical COFEC System Master Code List[4]

(Reprinted from Reliability Handbook *with permission of McGraw-Hill Book Company, Inc.)*

The fixed-format extraction routines of the report generator program constitute a single computer program that prepares all requested reports. Some combinations of reports can be prepared simultaneously. Other combinations require separate computer runs. The program might work in the following manner. A report request is read into the computer to establish which report or combination of reports is to be prepared, and designates an output file for each of the requested reports. Information from the data bank tape might then be stored in the core. Next, extraction routines corresponding to each of the requested reports are called, and the parameters needed in each report are extracted, tabulated where specified, placed in the requested format, and written on the designated file. At the appropriate interval, totals of the parameter values which have been accumulated are printed. When all records from the data bank have been processed, the individual output files are transferred to a single file to facilitate printing.

5-2.5 OPERATION OF THE DATA BANK

Operation of the data bank involves (1) its establishment and maintenance, and (2) its use as a source of parameters for reports. Its successful operation requires close coordination between engineering and data processing personnel. On large programs, a data coordinator and an engineering coordinator for data bank operations should be appointed to the program manager's staff. The details of establishing and maintaining the data bank and of extracting data from it are described in the paragraphs that follow (Ref. 7).

The operation of the data bank requires that the data be initially entered and then updated later. Coordinated effort between engineering and data processing personnel is required. The details of this operation are:

1. The request that information initially be entered into the data bank should be made by the program manager, to whom the equipment engineers respond by submitting completed data input forms certified by their signatures.

2. Revisions of data in the data bank are initiated by the responsible equipment engineer by requesting the necessary forms from the engineering coordinator, and submitting the completed and certified forms to him.

3. Completed data forms then should be reviewed technically by the program manager's staff. The data coordinator then reviews the forms for proper addressing of data, correct format, etc., and forwards the forms for data entry. The data forms then are returned to the data coordinator for checking against the revision report.

4. For the initiating operation, the EDP facility creates a new data bank file from all entries and issues a data bank file listing. For updating, the EDP facility merges data from the new entries with the old data bank file and creates a revision report listing all data additions, deletions, and revisions. The data bank file listing or the revision report is checked against the corresponding data forms by the data coordinator and submitted to the program manager for review.

Information is extracted from the data bank in the following manner:

1. The engineering coordinator initiates the request for a specific report to the data coordinator.

2. The data coordinator forwards a request form to the EDP facility, indicating the extraction routines to be used with the new data bank file, along with any special instructions.

3. The printed report then is submitted to the program manager by the EDP facility via the data coordinator.

A feature of the preceding operation is that all communication with the EDP facility is accomplished by the data coordinator. This procedure frees the equipment engineers and the program manager from all liaison with the facility.

Many newer programs will have online timeshared facilities that eliminate the need for much of the communication and red tape described in this paragraph.

5-3 RELIABILITY DATA SYSTEMS OPERATING PROCEDURES

The development of a reliability data system requires the creation of procedures, instruction, and forms for reporting, handling, and monitoring test data. It also requires computer programs for efficiently processing the data into formats suitable for reliability analyses.

Two important functions of the system operation will be discussed in detail in this paragraph: (1) data reporting, which establishes the requirements and instructions to be used by test personnel in reporting data; and (2) data control, which consists of monitoring the reported data to ensure its completeness, accuracy, and validity, and preparing it for computer input. Another important function, reliability-data reporting, will be discussed in par. 5-4.

In order to prepare an integrated reliability data system for a specific system, the system weapon specifications and technical development plan must be evaluated. As a result of these analyses, information derived from specific test data reporting requirements, the subsystem reliability model, and computer report requirements can be developed into the specific data reporting, control, and processing instructions required for the system.

5-3.1 DATA REPORTING

Data reporting procedures and instructions must be issued to the test groups for recording the test and failure data needed for reliability measurement and other purposes (Refs. 5 and 6). These procedures define the types of data to be reported, the forms to be used, and the instructions for completing them.

The data to be reported for reliability assessment should be defined in a standard data reporting requirements document. This document should be prepared as part of the contractor's reliability test program.

A data reporting group should be organized to develop the forms for data reporting and the procedures for their use. The forms that are established should provide instructions for the collection of the data required for reliability measurement. They also may contain information required for purposes other than reliability measurement.

The test information needed for reliability purposes includes:

1. Test Description
   a. Test report number
   b. Test level – component, equipment, or subsystem
   c. Test' type–qualification, acceptance, field, etc.
   d. Test site
   e. Test environment
   f. Date of test
   g. Test condition–operating, non-operating, or cycling.
2. Hardware Identification:
   a. Hardware name
   b. Hardware drawing number
   c. Hardware serial number
   d. Subhardware actually involved in test
   e. Hardware level
   f. Vendor.
3. Test Results:
   a. Time or cycles to failure
   b. Component failing
   c. Failure modes.

Accurate recording of failure data is essential to any reliability data system. Failure forms must be provided to test and operating personnel. When a failure occurs, a failure report – which contains a description of the failure, hardware identification, test conditions at the time of failure, cause of failure, and hardware disposition – must be written. If the failure occurred during test, the test document number must be referenced on the failure report. Other items of information are added to this report by the failure analysis system for processing purposes. This additional information usually consists of failure classification, classification as to relevancy, fault isolation code (used to identify the failed subcomponent in those components which are multifunctional), and the simulated mission environment that caused the failure to occur.

Problems often arise when data must be reported by field maintenance and logistic personnel who often are so busy at their own tasks that they do not have the time to record data properly. It may be necessary, in such a case, to assign reliability field personnel to the task of gathering data and filling out forms. This may be the biggest difficulty of all, in getting good field data.

Different forms should be developed for in-house and field use. The in-house forms should be used to record defects discovered during receiving inspection and acceptance testing. Additional in-house forms must be developed for use during reliability demonstration testing. Field forms should be designed which facilitate the easy reporting of failures occurring in the field.

A wide variety of forms has been developed. Several typical forms are presented in **Figs.** 5-3 to *5-5;* Chapter 9, Ref. 4, shows other forms. The data forms should be set up so that data entry can proceed directly.

5-3.2 DATA CONTROL

Careful control must be exercised over the data recording operations in order to assure timely and accurate data reporting. Procedures must be established for (1) collecting, reproducing, distributing, and filing test and failure forms; (2) handling the reported data; (3) methods of monitoring the data for compliance with requirements; (4) preparing the data for conversion into a medium acceptable to data processing; and (5) providing for corrections to the reported data. See Fig. 5-6 for a typical data processing system. Since the data being collected are for a computerized data processing system, processing instructions must be developed and the data must be tabulated in a format that permits automatic error checking

The paragraphs that follow are brief descriptions of the tasks involved for each of the preceding functions and responsibilities.

5-3.3 DATA HANDLING

Test data forms generated in the test areas (at all levels of test) should be sent to the data control area for processing and analysis. Failure data forms should be sent to the failure analysis group for review and classification.

When blank forms are sent to the test areas, they should contain whatever preprinted information – such as hardware identification – is feasible. This will reduce recording errors and save time for the test personnel. If a mechanized data processing system is being used, certain fields of data can be coded for ease of processing and minimizing transcription errors.

Procedures for data reporting should be written and standardized. These procedures should contain instructions as to when

1. REPORT NO. 42169 | 2. INITIAL REPORT NO. | 3. REPORTING ACTIVITY | 4. MISSILE TMS | 5. MISSILE SERIAL NO.

6. FAILED ITEM PART NO. | 7. FAILED ITEM S/N | 8. FAILED ITEM NAME | 9. FAILED ITEM MFR. | 10. F/I REF. DESIG.

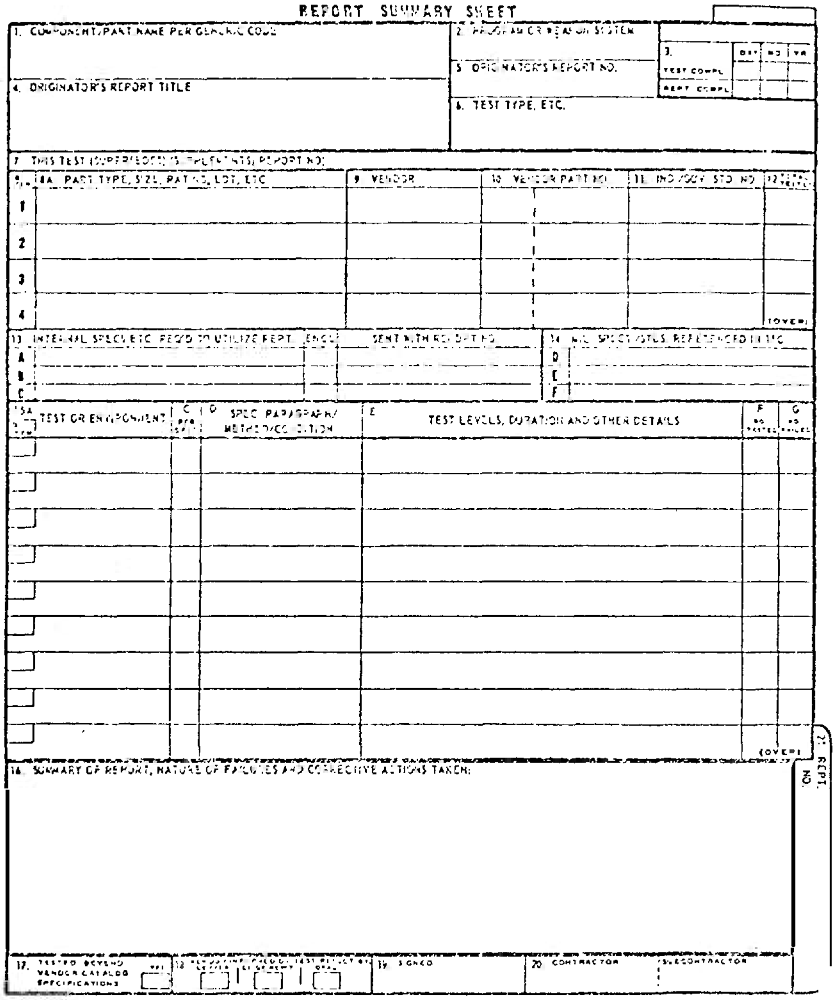
11. NEXT ASSY PART NO. | 12. NEXT ASSY NAME | 13. NEXT ASSY MFR. | 14. NEXT ASSY REF. DES. | 15. SYSTEM NO.

22. FAILURE DISCOVERED DURING	23. REASON FOR REPORT	24. REPAIR OR DISPOSITION ACTION	25. REPLACEMENT
.1 BENCH TEST .5 CHECKOUT .2 INSPECTION .6 MAINTENANCE .3 YTCRACE .7 MFR. TEST .4 SHIPPING .8 OPERATION	.1 FAILED ITEM .2 T.O. DIRECT .3 TIME EXPIRED .4 OTHER	.1 REPAIRED IN PLACE .5 CONDEMNED .2 REP REINSTALLED .6 HELD FOR REP. .3 ADJUSTED .7 DEPOT REP. .4 ELIMINATED .8 FAILURE ANALYSIS	.1 IDENTICAL PART .2 SUBSTITUTE PART .3 NONE NEEDED .4 NOT AVAILABLE

26.3 TEST CONDITION CODE | 26.4 ENVIRONMENT CODE | 26.5 SYSTEM AFFECTED | 27. REPORTED BY

***Figure 5-3. Typical Sample Failure Report*[4]**

***(Reprinted from* Reliability Handbook *with permission* of**
McGraw-Hill Book Company, Inc.)



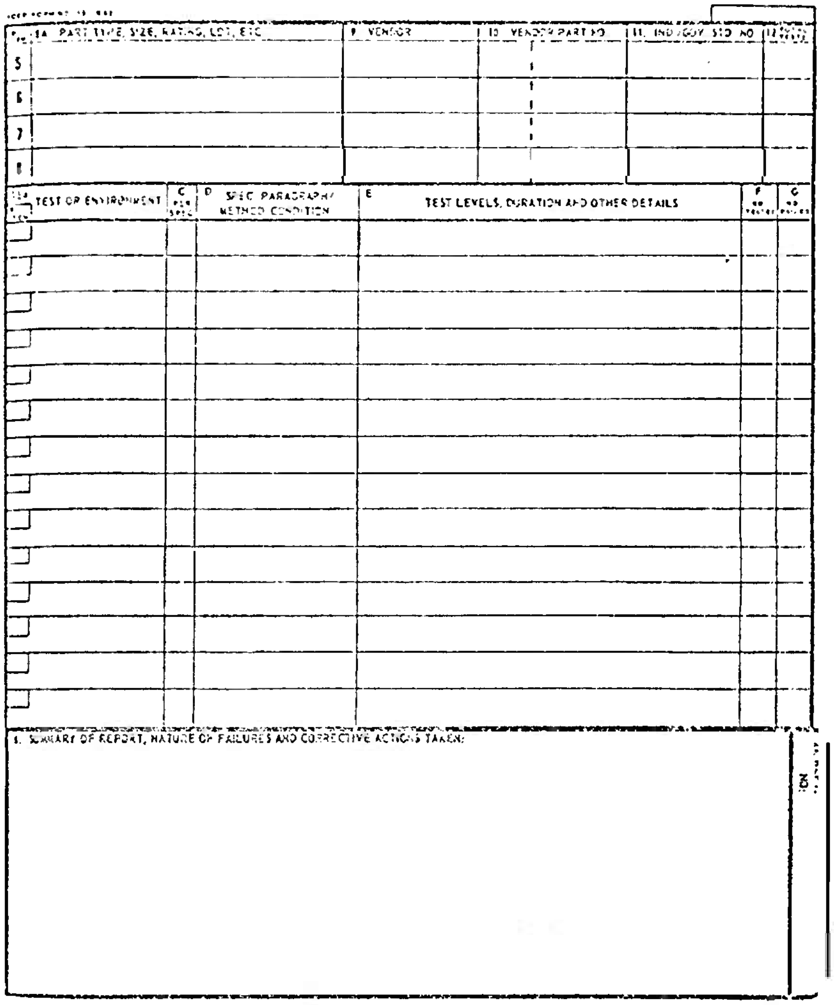
(A) Front of Form

Figure 5-4. Report Summary Format Stipulated in MIL-STD-831[5]

data will be reported, on what form, what disposition will be made of data sheets, as well as defining fields on the form and acceptable ranges for the data.

Personnel testing the hardware must accurately record all of the information required on the test forms and faithfully proceed through all steps in the program. For this reason, a training program should be established for them. Test personnel should attend periodic training programs to review the reporting forms and to discuss the proper information to be inserted into the various blocks on the forms.

Proper training is required for all personnel. However, personnel who are responsible for gathering data in the field or in other nontest environments should receive especially careful training in order to ensure that they properly record all data.



(B) Rear of Form

Figure 5-4. Report Summary Format Stipulated in MIL-STD-831[5]

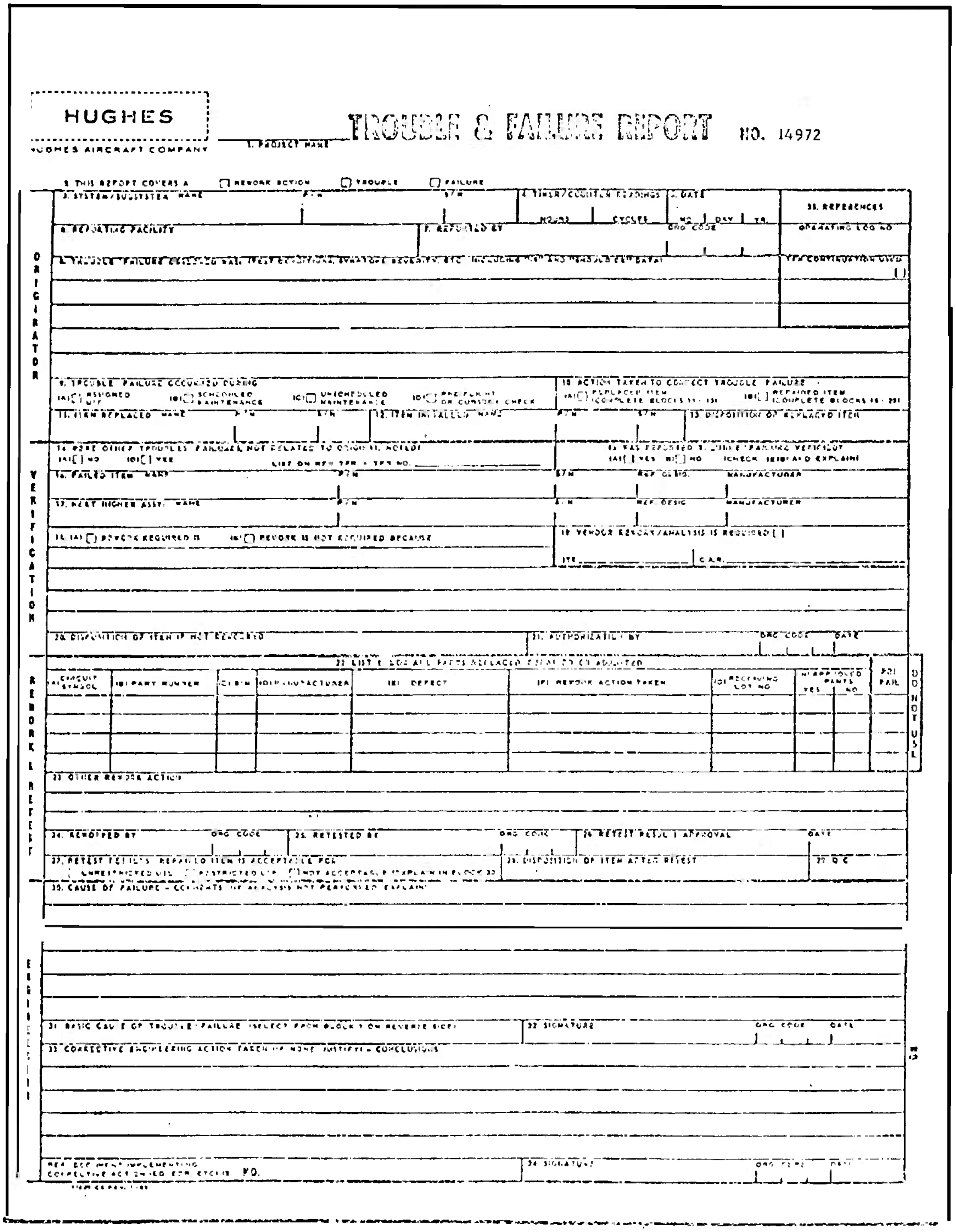
After all processing has been completed, the original data sheets should be filed for future reference.

5-3.4 DATA MONITORING

Army personnel responsible for monitoring the collection of data must have access to the test areas to assure that the tests are being run properly and that data are being recorded accurately.

After report forms are received by test personnel, they must be checked for gross errors – e.g., hardware identification, proper recording of test results, sign-off signatures, and legibility. When such errors occur in the test or failure documents, they must be returned to the originator for correction.

One method of assuring that all test data are being collected and processed is to use


HUGHES

HUGHES AIRCRAFT COMPANY

1. PROJECT NAME

TROUBLE & FAILURE REPORT

NO. 14972

(A) Page 1

Figure 5-5. Hughes Trouble & Failure Report[6]

HUGHES

TROUBLE & FAILURE REPORT

NO. 14972

CLOSEOUT SHEET

THE REPORTING REQUIREMENTS OF GOVERNMENT SPECIFICATIONS ARE FULLY SATISFIED WHEN ALL SECTIONS OF THE TFR ARE COMPLETED

GENERAL INSTRUCTIONS

1. THE TFR IS USED TO DOCUMENT THE TROUBLE FAILURE OF ONLY ONE (1) ITEM PER REPORT. IF MORE THAN ONE ITEM IS INVOLVED IN A SINGLE FAILURE EVENT, OR IF TWO OR MORE FAILURES ARE DISCOVERED IN A SINGLE ITEM, A SEPARATE REPORT SHALL BE COMPLETED FOR EACH ITEM OR FAILURE AND IDENTIFIED BY CROSS REFERENCING IN BLOCK 35
2. UNLESS INSTRUCTED OTHERWISE, ALL REQUIRED BLOCKS MUST BE COMPLETED
3. USE A TFR CONTINUATION SHEET WHENEVER ADDITIONAL SPACE FOR ENTRIES IS REQUIRED
4. INSURE LEGIBILITY OF LAST COPY

DETAIL INSTRUCTIONS

ENTRIES ON THE TFR ARE SELF-EXPLANATORY AND DO NOT REQUIRE DETAIL INSTRUCTIONS. SPECIAL INSTRUCTIONS FOR ENTRIES ARE COVERED IN IMPLEMENTING INSTRUCTIONS ISSUED BY THE USING ORGANIZATION

RELIABILITY ANALYSIS AND FOLLOW-UP

1. CAUSE OF DIFFICULTY

- [ ] DESIGN
- [ ] MFG. PROCESS
- [ ] OPERATING TIME
- [ ] TEST EQUIPMENT
- [ ] TEST PROCEDURE
- [ ] HUMAN ERROR
- [ ] WORKMANSHIP
- [ ] WIRING ERROR
- [ ] ROUGH HANDLING
- [ ] ENVIRONMENTAL EFFECT(S)
- [ ] OTHER ITEM FAILURE ________ SPECIFY
- [ ] DEFECTIVE PARTS
- [ ] UNKNOWN
- [ ] OTHER ________ SPECIFY

2. EFFECT ON

________ (ITEM NAME)

- [ ] INOPERATIVE
- [ ] BADLY DEGRADED
- [ ] SLIGHTLY DEGRADED
- [ ] NUISANCE

3. CORRECTIVE ACTION STATUS — DATE

- [ ] IN EFFECT ________
- [ ] UNDER INVESTIGATION ________
- [ ] SCHEDULED ________
- [ ] INTERIM FIX ________
- [ ] NOT REQUIRED ________
- [ ] ACTION PREVIOUSLY TAKEN ________

C. A. R. NO. ________

4. REPLACED/REPAIRED ITEM OPERATIONAL USAGE

HOURS | CYCLES

5. RELATED CASES:

6. CLASS OF FAILURE

PARTS ANALYSIS

7. CIRCUIT SYMBOL	8. RESULTS	9. ANALYZED BY	ORG CODE	DATE	10. FAILURE CODE

11. COMMENTS

12. RELIABILITY ANALYSIS BY ________ ORG. CODE ________ DATE ________

APPROVALS:

MO. | DAY | YR.

13. TFR CLOSED

________ VENDOR/CUSTOMER (IF APPLICABLE) ________ ENGINEERING (IF APPLICABLE) ________ RELIABILITY

10396 SS 10-65

(B) Page 2

Figure 5-5. Hughes Trouble & Failure Report[6] (Cont'd)

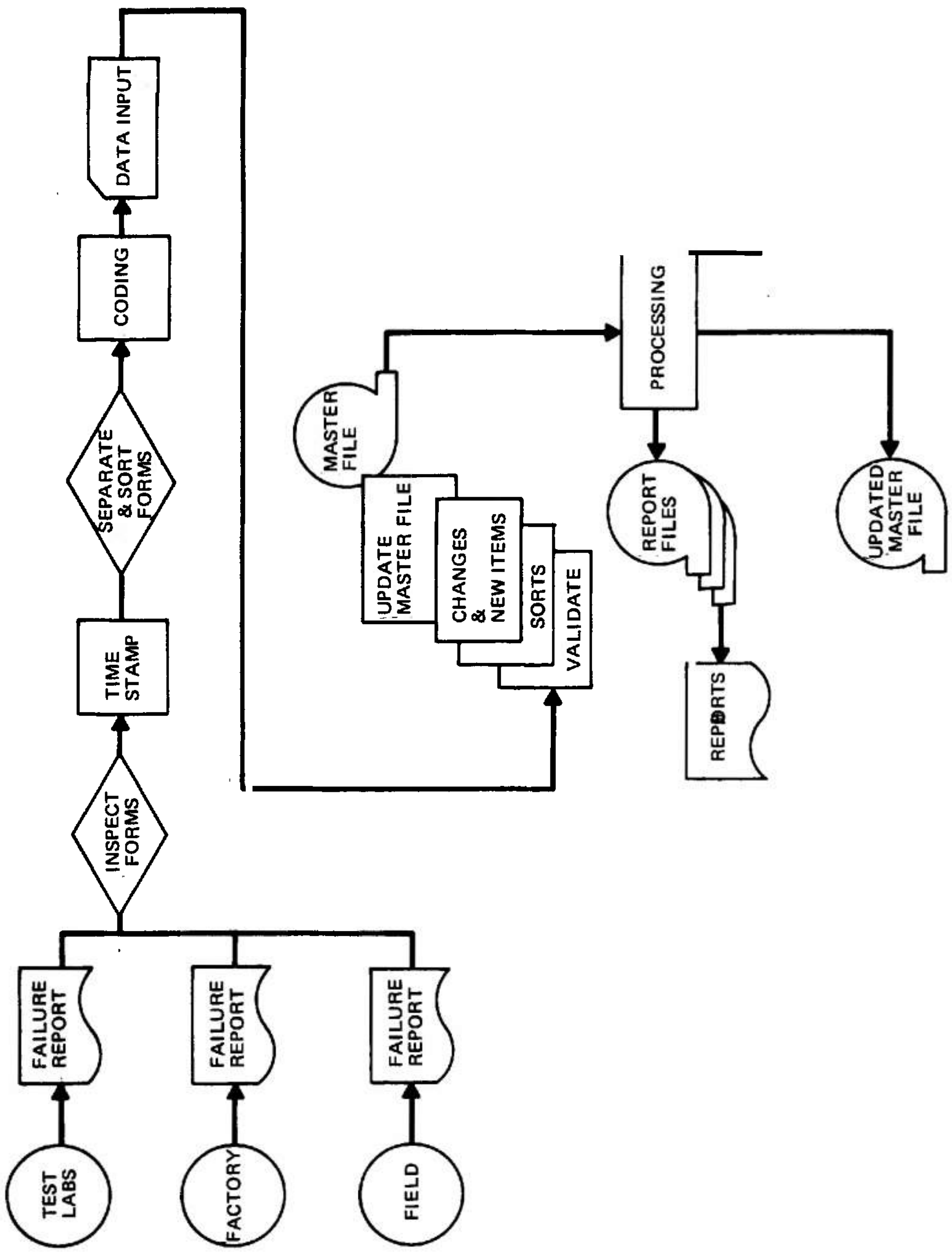


Figure 5-6. Typical Data Processing System

a check list containing the drawing numbers and serial numbers of items scheduled to be tested, and which can be compared against the data being processed. Check list reports are returned to data control if data are missing.

5-3.5 PREPARING TEST DATA FOR COMPUTER PROCESSING

The information contained in test and failure documents must be processed to assure that the input data records are in proper format for computer usage.

There are two basic types of input data: test information and failure information. These usually exist in separate documents. This segregation of data is carried through input processing. The data are submitted for data processing in that segregated fashion.

The instructions for processing the input data must contain detailed information on:

1. The location of data fields on original test or failure documents

2. The placement of data fields on the input records

3. Special handling that may be required for any particular data form or field of information.

After input processing has been accomplished, the input records (cards, tape) should be forwarded to data processing for entry into the processing system.

5-3.6 ERROR CORRECTION

When errors are detected in the data during error checking, listings explicitly defining the errors should be sent to the data control group. Data control personnel then must correct the errors and return the data sheet to the area responsible for that error. The corrected data are prepared for processing and reinserted into the processing cycle.

5-4 RELIABILITY REPORTING

The contractor's reliability reporting group should prepare all necessary in-house and contractual reports for the reliability test and measurement program. The general contents and procedures for generating these reports must be developed in accordance with Military Specifications and the particular contract. These procedures and the report contents are a function of the specific subsystem reliability reporting requirements.

The following reports should be prepared:

1. Reliability Status

2. Failure Summary

3. Historic Test Result

4. Failure Status

5. Hardware Summary

6. Failure Analysis Follow-up

7. Failure Rate Compendium.

The reports described in this paragraph can be used by the contractor in-house as well as by the Army. The contractor can use them for purposes such as estimating spares and logistic requirements, maintaining a test history by serial number of critical and limited life items, or for establishing a failure rate compendium based on actual test and field experience. The status reports produced by the contractor for the Army can vary from project to project.

The contents and format of reliability status reports, which must be issued periodically throughout the program, should be discussed by the Army and the contractor to arrive at a mutually agreeable document. In many cases, reliability reports generated by the contractor's normal procedures can be used by the Army without modification.

5-4.1 RELIABILITY STATUS REPORTS

The reliability status report presents estimates of the reliability of each component, equipment, and subsystem; pinpoints reliability problem areas; and discusses possible corrective actions.

The reliability status report should contain:

1. A brief description of the subsystem and equipment operation, and mission against which reliability is reported

2. The subsystem block diagram and reliability equation, or an equivalent fault tree or cause-consequence chart

3. A summary of the sources of test data

4. A table relating the current measured reliability and reliability requirements for each hardware level

5. Growth curves that tabulate measured reliability versus time

6. A discussion of reliability problem areas, proposed corrective action, and the results of previous corrective action

7. A tabulation of the failure rates and reliability estimates at the component, equipment, and subsystem levels (see Fig. 5-7)

8. The s-confidence levels at which the required reliability was demonstrated by each hardware element.

The composite reliability status report in Fig. 5-8 and the reliability status report supplement in Fig. 5-9 represent two useful formats for summarizing reliability status. In the fiist report, failure rates, reliability indices, s-confidence levels, and mission information are summarized by environmental test category for each hardware item tested.

5-4.2 FAILURE SUMMARY REPORTS

The failure summary report presents a complete record of all failures occurring on a particular program. Making this report mandatory helps to ensure that failures are identified and reported properly, that important and repetitive failures are analyzed in detail, that causes and modes of equipment failures are determined, and that corrective actions are developed.

Monthly failure summary reports should contain the following information:

1. Hardware identification – including nomenclature, drawing number, serial number, vendors, and program code

2. Test description – including test type, environment, site or reporting activity, and date of test

3. Test results – including failure report number, failure classification, and description of failure

4. Failure investigation analysis, including the corrective action recommended or taken

5. Names of the responsible personnel.

A sample failure summary report is shown in Fig. 5-10.

EQUIPMENT RELIABILITY STATUS REPORT

PROJECT A3 C1 • 50% CONFIDENCE (BEST ESTIMATE) C2 • 80% CONFIDENCE CURRENT DATE PAGE #

NOMENCLATURE	NUMBER	MISSION A				MISSION B				MISSION C			
				RELIABILITY		FAILURE RATE		RELIABILITY		FAILURE RATE		RELIABILITY	
		λ_{C1}	λ_{C2}	R_{C1}	R_{C2}	λ_{C1}	λ_{C2}	R_{C1}	R_{C2}	λ_{C1}	λ_{C2}	R_{C1}	R_{C2}
RECONNAISSANCE VEHICLE SUBSYSTEM		.0433806	.0755389	9575	.9273	0230083	.0418066	.9773	.9591	.0827581	.1441077	9205	.865
TELEMENTRY & COMMAND EQUIPMENT													
TRANSOUTOR	0194C0543G003	.0038418	.0129154	9962	.9872	0034575	.0116207	.9965	.9884	.0061469	.0206359	9939	.9796
TRANSOUTOR	0194C0543G003	.0038418	.0129154	9962	.9872	0034576	.0116207	.9965	.9884	.0061469	.0206359	9939	.9796
PROCRAMMER	0194C0696G001	.0028312	.0116673	9952	.9884	0017363	.0103553	.9983	.9897	.0039563	.0186699	9961	.9815
AMPLIFIER	0604D0147G001	.0033467	.0125532	9967	.9875	0020166	.0113457	.9980	.9887	.0057011	.0199872	9943	.9803
MULTICODER	0926B0972G001	.0000000	.0011743	9999	.9988	0000000	.0006572	.9999	.9993	.0000000	.0207191	9999	.9795
SENSORS			.0007143	9996	.9993	0003993	.0007512	.9996	.9992	.0083293	.0175541	9917	.9826
BAROSWITCH			.0200054	9890	.9801	0060607	.0151016	.9994	.9850	.0150113	.0289132	9851	.9715
TELEMETRY & COMMAND EQUIPMENT GROUPING		.0252424	.0615249	9751	.9404	0136711	.0333214	.9863	.9672	.0391449	.0954103	9617	.9090
COMMUNICATIONS EQUIPMENT													
BEACON	0563E0617G001	.0004861	.0016290	9935	.9984	0003579	.0007145	.9996	.9993	.0015500	.0037440	9984	.9963
PULSE GENERATOR	0215E0175G001	.0034357	.0113543	9966	.9876	0005678	.0095312	.9995	.9905	.0110345	.0190534	9890	.9810
FLASHISC LIGHT	0681C0643G003	.0000000	.0007365	9999	.9993	0000000	.0004932	.9999	.9995	.0000000	.0011223	9999	.998
RECEIVER	0194C0796P001	.0009895	.0013563	9991	.9987	0006563	.0010976	.9993	.9989	.0012123	.0012013	9988	.995
TRASSMITTER	0926B0979P001	.0027573	.0088379	9973	.9912	0051234	.0065374	.9950	.9935	.0102567	.0122667	9890	.9878
COMMUNICATIONS EQUIPMENT GROUPING		.0076686	.0153975	9923	.9847	0067454	.0135438	.9933	.9864	.0240535	.0482961	9762	.952
VEHICLE EQUIPMENT													
VEHICLE COSTROLLER	0179D0869G003	.0036998	.0101933	9963	.9899	0025001	.0084567	.9975	.9915	.0101023	.0189934	9899	.981
PROPULSION	0215E0168G002	.0060709	.0094637	9939	.9905	0007011	.0076776	.9993	.9923	.0084567	.0136245	9915	.986
POWER SUPPLY	0179D0870P002	.0006989	.0011097	9993	.9988	0003911	.0010594	.9996	.9989	.0010010	.0036364	9990	.996
VEHICLE EQUIPMENT GROUPING		.0104696	.0164005	.9895	.9837	0035923	.0056271	.9964	.9944	.0195600	.0306390	.9806	.969
PAYLOAD EQUIPMENT													
TV CAMERA	0215E0133G001	.0000000	.0017563	9999	.9983	0000000	.0009911	.9999	.9991	.0000000	.0025394	9999	.997
PAYLOAD EQUIPMENT GROUPING		.0000000	.0017563	9999	.9983	0000000	.0009911	.9999	.9991	.0000000	.0025394	9999	.997

Figure 5-7. Table from a Sample Reliability Status Report[7]

5-4.3 HISTORIC TEST RESULT REPORTS

A historic test result file, containing records of all reliability tests, is maintained in the reliability data bank. The data in this file are used to produce the historic test result report, an example of which is shown in Fig. 5-11. The report should contain, as a minimum, the following information:

1. Hardware Identification

    a. Name

    b. Drawing number

    c. Contractor serial number

    d. Vendor serial number

    e. Vendor identification (Federal Handbook Code)

COMPOSITE RELIABILITY STATUS REPORT

PROJECT A3 COMPONENT LEVEL REPORT C1 = 50% CONFIDENCE (BESTESTIMATE) C2 = 80% CONFIDENCE CURRENT DATE PAGE #

DRAWING NUMBER 0104C0543C003 NOMENCLATURE TRANSOLATOR TEST TYPES

ENVIRONMENT	COSDITION A HOURS	MIHS	N	F	CONDITION B HOURS	MINS	N	F	COSDITION C CYCLES	N	F	CONDITION D HOURS	MISS	N	F	TOTAL U	F	QUAL U	F	ACCEPT U	F	FIELD U	F
TEMPERATURE	429	46	53	1								57	15	53		50	1	11	1	39			
VIDRATJON	20	04	52															11		39			
LIFE												635	16	156		52		11		41		13	

	EQUIVALENT MISSIONS COSD. A	COSD. B	COND. C	COSD. D	NORMALIZING ALPHA VALUES COND. A	COND. B	COND. C	COND. D
MISSION A								
TEMPERATURE	257.86			68. 7	1/100			1/50
VIBRATION	140.333				1/12			
LIFE				211,755				1/180
MISSION 5								
TEMPERATURE	236.511			76.3333	1/90			1/45
VIBRATION	336.8				1/05			
LIFE				254. 106				1/80
MISSION C				42.9375	1/160			1/00
TEMPERATURE	161,162				1/40 .			
VIBRATION	42. 1			159.816				1/240
LIFE								

	FAILURE RATES C1	C2	RELIABILITY INDICES C1	C2	nC	n	λC
MISSION A							
TEMPERATURE	.0038418	.0129154	.9962	.9872	.014556	1.0	.000014898
VIBRATION	.0000000	.0117112	.9999	.9883	.007126	1.0	.0
LIFE	.0000000	.0077623	.9999	.9922	.004723	1.0	.0
MISSION TOTAL	.0038418	.0129154	.9962	.9872	.026405	3.0	.000014898
MISSION B							
TEMPERATURE	.0034576	.0116207	.9965	.9884	.013100	1.0	.000012068
VIBRATION	.0000000	.0048798	.9999	.9951	.002969	1.0	.0
LIFE	.0000000	.0064677	.9999	.9935	.003935	1.0	.0
MISSION TOTAL	.0034576	.0116207	.9965	.9884	.020004	3.0	.000012068
MISSION C							
TEMPERATURE	.0061469	.0206359	.9939	.9796	.023289	1.0	.000038140
VIBRATION	.0000000	.0390343	.9999	.9617	.023750	1.0	.0
LIFE	.0000000	.0103482	.9999	.9897	.006296	1.0	.0
MISSION TOTAL	.0061469	.0206359	.9939	.9796	.053335	3.0	.000038140

PROJECT RESULTS

C1	C2	C1	C2	REQUIRED RELIABILITY	DEMONSTRATED CONFIDEKCE
.0134463	.0284202	.9866	.9719	.9899	.32

LEGEND: n IS NUMBER OF TERMS IN C CALCULATION

Figure 5-8. Sample Composite Reliability Status Report'

PAGE f

EQUIPMENT RELIABILITY STATUS REPORT SUPPLEMENT

PROJECT A3 U = UNITS TESTED F = NUMBER OF FAILURES C1 = 50% CONFIDENCE (BEST ESTIMATE) C2 = 80% CONFIDENCE 'CURRENT DATE

NOMENCLATURE	DRAWING NUMBER	ENVIRONMENT	TEST CLASSIFICATION						FAILURE RATE		RELIABILITY	
			QUAL		ACCEPT		FIELD					
			U	F	U	F	U	F	λ_{C1}	λ_{C2}	R_{C1}	R_{C2}
RECONNAISSANCE VEHICLE SUBSYSTEM												
TELEMETRY - COMMAS- EQUIPMENT												
TRANSOLATOR	0194C0543G003	MISSION A										
		TEMPERATURE	11	1	39				0036418	.0129154	.9962	.9872
		VIBRATION	11		39				0000000	.0117112	.9999	.9583
		LIFE	11		41		13		0000000	.0077623	.9999	.9922
		MISSION B										
		TEMPERATURE	11	1	39				0034576	.0116207	.9965	.9884
		VIBRATION	11		39				0000000	.0048799	.9999	.9951
		LIFE	11		41		13		0000000	.0064677	,9993	.9935
		MISSION C										
		TEMPERATURE	11	1	39				0061469	.0206359	.9939	.9796
		VIBRATION	11		39				0000000	.0390343	.9999	.9517
		LIFE	11		41		13		0000000	.0103482	.9999	1.9997
PROGRAMMER	0194C0696G0001	MISSION A										
		TEMPERATURE	6		40				0000000			.9987
		VIBRATION	6		40	1			00283 IS			.9984
		LIFE	6		39		10		0000000			.9835
		MISSION B										
		TEMPERATURE	6		40				0000000			.9921
		VIBRATION	6		40	1			5017369			,9997
		LIFE	6		39				0030000			.9315
		MISSION C										
		TEMPERATURE	6		40				0C00000			.9951
		VIBRATION	6		40	1			0039563			.9915
		LIFE	6		39		10		0000030			,9908
AMPLIFIER	0604D0147G001	MISSION A										
		TEMPERATURE	6		39				0000000			.9917
		VIBRATION	6		39				0000000			.9899
		LIFE	6		39	1	12		0033467			,9675
		MISSION B										
		TEMPERATURE	8		39				0000000	.0095578	.9999	.9905
		VIBRATION	6		39				0000000	.0289122	.9999	.9715
		LIFE	6		39	1	12		0020156			.9887
		MISSION C										
		TEMPERATURE	6		39				0000000			.9888
		VIBRATION	6		39				0000000			.9815
		LIFE	6		39	1	12		0057011			.9803

Figure 5-9. Sample Reliability Status Report Supplement[7]

f. Project code

g. Hardware level.

2. Test Description:

a. Hardware level

b. Test type

c. Environment

d. Site

e. Date of test

f. Test report number.

3. Test Results:

a. Test time

b. Test cycles

c. Test failures

d. Failure report numbers

FAILURE SUMMARY REPORT

	D.	DRAWING NUMBER 103 C 4277 GI
MANUFACTURER 04615	O. C. ENGINEER W. E. Build	EQUIPMENT NAME VHF Transmitter

FAIL RPT. NO.	RE DATE	RPTG ACTVTY	SERIAL NO.	TEST TYPE	TEST ENVIRONMENT	FAILURE DESCRIPTION	FAILURE CLASS	FAILURE REQ	FAILURE INVESTIGATION/ANALYSIS CONCLUSION	CORRECTIVE ACTION RECOMMENDED/TAKEN
39138	10/7/6	Phila.	1	Qual.	Humidity	Frequency out of spec.; moisture entered transmitter due to faulty test assembly	30000	Yes	Human error, assembly screws not properly tightened against component itself, when not attached to base plate. (F. A. R. A-678)	R: Issue AN and SI to eliminate the possibility of recurrence. T: AN 688 E 585-5 and TR 8047-2 to SI 21834
39138	10/7/6	Phila.	2	Qual.	Humidity	Frequency out of spec.; moisture entered transmitter due to faulty test assembly	30000	Yes	Human error, assembly screws not properly tightened against component itself, when not attached to base plate. (F. A. R. A-678)	R: Issue AN and SI to eliminate the possibility of recurrence T: AN 688 E 585-5 and TR 8047-2 to SI 24834
						20 November 1961				
352-70	3/14/6:	AMR	5476204	Pre-launch	Ambient	Emitting sidebands & noise equal in amplitude to main carrier. Signal strength very low	33030	Yes	Improper Tuning due to lack of adequate tuning procedure. (F. A. R. A-772)	R: Develop tuning procedure and instruct personnel in use T: Procedures demonstrated and distributed to field personnel.
352-11	12/26/61	AMR	5476043	Hangar	Ambient	Low Power output.	01000	Yes	Discrepancy between System and Component requirements. (F.A.R. A-772)	R: Change system spec. to conform to component requirements. T: Unable to secure permission to change spec. to 9 watt minimum. Part selection still done for 10 watt minimum.
						20 March 1962				
31176	6/7/62	Phila.	N/A	Syst. O/A	Ambient	Low power output.	31000	NO	Cable to power amplifier too long.	
						30 June 1362.				
49095	5/31/6:	Phila.	Lot 11-I	Comp. O/A	Post Vibration	Power drops intermittently.	02020	No	Defective insulator on Q4 heat sink.	T Replaced insulator.
						20 August 1962				
86720	8/22/6:	Phila.	Lot 9-2	Comp. O/A	Vibration	Chassis shorted to case	01000	NO	Insulation shorted.	T: Removed from case and replaced shorted insulator.
AC-2257	8/24/61	AMR	54'76572	Pre-launch	Ambient	Multiple oscillation above 9.5 watts	01000	NO		
						20 September 1962				
02433	1/9/63	Phila.	A30	Comp. O/A	Vibration	Unit broke into oscillation, power output and input current dropped.	01000	No	Defective diode, CR6.	T Replaced
• 02436	1/9/63	Phila.	A27	Comp. O/A	Vibration	No output, no oscillations from oscillator.	01000	No	Defective transistors Q1, 7, 8, 9.	T Translators replaced.
						20 March 1963				

Figure 5-70. Sample Failure Summary Report'

PROJECT A3 HISTORIC TEST RESULTS FILE CURRENT DATE PACE #

	REC LEV	TEST DESCRIPTION: TYPE	ENV	SITE	DATE: YR	MO	DY	LEV	TEST CONDITIONS: HRS	MN	F	HRS	MN	F	CYCLES	F	HRS	MS	F.	TEST DOC NO.	FAIL DOC SO.	FLR CLS	ELM CLS	SHOP ORDR	ENTRY DATE	REFERESCED: TEST DOC NO.	FAIL DOC NO.	PROJECT
NAME TRANSOLATOR		DWG. SO. 0194C0543G003							COST. SER. #5227766																	VENDOR CODE 04615		
	C	OA	TM	DP	63	12	01	C	8	30							1	20										
	C	OA	VB	DP	63	12	01	C		43																		
	C	OA	ZA	DP	63	12	02	C									12	10										
	C	OA	ZB	DP	63	12	01	C									1											
	C	OA	ZC	DD	63	12	02	C										30										
NAME TRASSOUTOR		DWC. NO. 0194C0543G003							COKT. SER. #5227767											VESDOR SER. 1122						VESDOR CODE 04615		
	C	04	TM	DP	63	12	02	C	8	30							1	20		DP1202				8432	120563			
	C	OA	Vn	DP	63	12	02	C		48										DP1202				8432	120563			
	C	OA	ZA	DP	63	12	02	C									12			DP1202				8132	120563			
	C	04	ZB	DP	63	12	02	C									1	05		DP1202				8432	120563			
	C	OA	ZC	DP	63	12	03	C										30		OP1202				8432	120563			
NAME TRANSOUTOR		DWG. SO. 0194C0543G003							COST. SER. 15227769											VESDOR SER. C123						VENDOR CODE 04615		
	C	QU	TM	DQ	63	12	05	C		45	1									DQ0974	39339	0	1	8432	121263			
	C	QU	TM	DQ	63	12	05	C	16	40	1							10		DQ0975	39344	2	0	8432	121263	DQ0974	39339	
	C	QU	TM	DQ	63	12	10	C	16	30							2	30		DQ0988				8432	122063	DQ0975		
	C	QU	VB	DQ	63	12	06	C		35										DQ0975				8432	121263	DQ0974		
	C	QU	ZA	DQ	63	12	04	C									22	15		DQ0975				8432	121263	DQ0974		
	C	QU	2A	DQ	63	12	09	C									22	15		DQ0988				8432	122063	DQ0975		
	C	QU	ZB	DQ	63	12	06	C									1	30		DQ0975				8432	121263	DQ0974		
	C	QU	ZB	DQ	63	12	11	C									1	30		DQ0988				8432	122063	DQ0975		
	C	QU	ZC	DQ	63	12	06	C										45		DQ0975				8432	121263	DQ0974		
NAME PROCRAMMER		DWG. SO. 0194C0696G001							COST. SER. #5227833											VENDOR SER # 2104						VENDOR CODE 04615		
	C	OA	TM	DP	63	01	27	C	2	45										DP1092				8432	020363			
		OA	VB	DP	63	01	28	C		50										DP1092				8432	020963			
		OA	ZA	DP	63	01	27	C							12					DP1092				8432	020863			
		OA	ZB	DP	63	01	27	C							6					DP1092				8432	020863			
	C	OA	ZC	DP	63	01	28	C							3					DP1092				8232	020863			
NAME PROGRAMMER		DWC. SO. 0194C0696G001							CONT. SER. ,5227834											VENDOR SER. #2105						VENDOR CODE 04615		
		OA	TM	DP	63	02	01	C	2	45										DP1093				8432	020363			
		OA	VB	DP	63	03	02	C		10	1									DP1093	39108	1	0	8432	020363			
		OA	VB	DP	63	02	03	C		50										DP1094				8432	020963	DP1093		
		OA	ZA	DP	63	02	01	C							12					DP1093				8432	020863			
		OA	ZB	DP	63	02	02	C							6					DP1093				8432	020863			
		OA	ZB	DP	63	02	03	C							6					DP1094				8432	020663	DP1093		
		OA	ZC	DP	63	02	03	C							3					DP1094				8432	020863	DP1093		
SAKE PROGRAMMER		DWG. SO. 0194C0690G001							CONT. SER. #5227835											VENDOR SER. #2106						VENDOR CODE 04615		
		OA	TM	DP	63	02	03	C	2	45										DP1095				8432	021563			
		OA	VB	DP	63	02	03	C		50										DP1095				8432	021563			

NOTE: CODED INFORMATION IS DEFINED IN TABLE 4-1.

LEGEND: HRS = HOURS; MN = MINUTES; F = NO. OF FAILURES

Figure 5-11. Sample Historic Jest Results File'

e. Failure classification

f. Fault isolation code (identifies hardware levels to which each failure can be attributed)

g. Timer number (identifies the time meter from which data were obtained).

4. Reference Information:

a. Date of entry (date information reaches file)

b. Test report number (references the test from which the record was generated)

c. Failure report number (references a higher level failure report)

d. Project codes (identifies data used from other programs).

The historic test result file permits a wide variety of reports to be prepared. For example, by using a control on date of test, reports containing data generated over any desired range of dates can be prepared. Use of a control on level of test permits reports containing only subsystem data or any desired combination of subsystem, equipment, or component data to be prepared. Use of a control on test type permits reports to be generated which eliminate engineering development and acceptance screening tests or which use only field data. A properly structured historic test result file permits reporting and analysis of any desired combination of the stored data.

5-4.4 FAILURE STATUS REPORTS

The purpose of the failure status report is to maintain an historical record of all failures. The report contains an entry for each failure. The equipment identification and test description from the historical file should be provided in this report along with those portions of the test results and referenced information that relate to failures (i.e., failure report number, failure classification, include-exclude criteria, fault isolation code, and referenced failure report numbers).

The failure status reports form the basis of the failure summary. A description of the failure, conclusions based on the analyses, and the corrective action recommended or taken should be added to the status report to form the summary. A sample sheet from a failure status report is shown in Fig. 5-12.

5-4.5 HARDWARE SUMMARIES

The hardware summary report should contain an entry for each hardware tested, consisting of the total test time and failures accumulated. This report can be used in logistical planning.

The hardware summary report can be expanded to include failure rates of items required for early spares provisioning estimates, if these items are not included in the reliability measurement plan.

5-4.6 FAILURE ANALYSIS FOLLOW-UP REPORTS

A failure analysis follow-up report should be issued periodically for internal action. This report should list every item which requires further action and its status.

5-4.7 FAILURE RATE COMPENDIA

Perhaps the most valuable byproduct of the reliability test program is the failure rate compendium which is a compilation and summary of the hardware test results contained on the historic test result file. The data from all projects should be summarized by hardware groupings to provide a refer-

PROJECT A3 FAILURE STATUS REPORT CURRENT DATE PAGE#

NOMENCIATURE	DRAWING NO.	FAIL DOC ND	REF DOC	FAIL NO.	CONT. SER. KO.	VENDOR SER. NO.	TYPE	TEST DESCRIPT ENV	SITE	DATE	N LEV	COND.	FLR CLS	ANAL REQ.	ELM CLS	PRJ CD	TEST DOC NO.	REF. P9J	SHOP ORDR	. I C
VEHICLE CONTROLLER	0179D0869G003	39333			5226785	347	OA	TM	DP	042863	C	A	3	KO	1	A3	DP0635		8432	
		39354		9333	5226735	347	OA	ZC	DP	062363	C	D	2	YES	1	A3	DP0789		8432	
		39401			5226453	288	OA	ZB	DP	013063	C	D	2	YES	0	A3	DP0595		8432	
POWER SUPPLY	0179D0870P002	39095			5325476	7390	QU	ZB	KA	122463	C	D	1	YES	0	A3	DP0983		8132	N12
		39387			5225593	7450	OA	ZA	DP	081263	C	D	2	YES	1	A3	DP0805		8432	
		39392			5225539	7462	OA	VB	DP	091363	C	A	1	YES	1	A3	DP0856		8432	
		39409			5225061	7466	Oh	VB	DP	101063	C	A	1	YES	1	A3	DP0901		8432	
TRASSOUTOR	0194C0543G003	39344		9339	5227768	123	QU	TM	DQ	120563	C	A	2	YES	0	A3	DQ0975		8432	
		39339			5227768	123	QU	TM	DQ	123563	C	A	0	NO	1	A3	DQ0974		8432	
PROGRAMMER	0194C0696G001	39108			5227834	2105	OA	VB	DP	320263	C	A	1	YES	0	A3	DP1093		8432	
RECEIVER	0194C0706P001	38098			5222223	3467	QU	ZA	DQ	100962	C	D	1	YES	0	A3	DQ0555		8432	
		38099			5222325	3469	QU	ZA	DQ	102362	C	D	2	YES	0	A3	DQ0567		a432	
		38110		8098	5222323	3467	QU	ZB	DQ	111562	C	D	3	YES	1	A3	DQ0601		8432	
		38150			5222326	3470	QU	VB	DQ	111662	C	D	3	NO	1	A3	DQ06 10		8432	
		38103			5222327	3471	QU	TM	DQ	112362	C	D	3	NO	0	A3	DQ0705		8432	
		38174			5222329	3473	OA	VB	DQ	030663	C	A	0	NO	1	A3	DP0563		8432	
TV CAMERA	0215E0133G001	39200			5236678	11	OA	TM	DP	121663	C	A	1	YES	0	A3	DP0998		8432	
PROPULSION	0215E0168G002	39210			5246711	4	OA	TM	DP	010364	C	A	1	YES	0	A3	DP1111		8432	
		39211		3210	6246711	4	OA	TM	DP	010364	C	A	3	NO	1	A3	DP1 112		8432	
		39212		3210	5246711	4	DA	VB	DP	010304	C	A	3	NO	1	A3	DP1112		8432	
PULSE GENERATOR	0215E0176G001	38660			522313	466	OA	TM	DP	060663	C	A	0	SO	1	A3	DP0876		8432	
		33667			522314	467	OA	ZA	D?	060653	C	C	0	SO	1	A3	DP0877		8432	
		39693			522393	546	OA	TM	DP	070363	C	A	0	SD	1	A3	DP0901		8432	
		38701		3667	522313	4F6	DA	ZB	DP	071063	C	C	3	YES	0	A3	DP1003		8432	
		38710			522367	520	OA	ZB	DP	072063	C	C	3	SO	0	A3	DP1005		8432	
		33733			522403	556	OA	VB	DP	072963	C	A	2	YES	0	A3	DP1015		8432	
		39610			522939	1152	OA	TM	DP	090163	C	A	3	YES	1	A3	DP1163		8432	
		39620			523010	1163	OA	ZA	DP	121363	C	C	3	NO	1	A3	DP1169		8432	
BEACON	0563E0617G001	38998			523669	3667	OA	TM	DP	101063	C	A	0	NO	1	A3	DP0963		8432	

NOTE: CODED INFORMATION IS DEFINED IN TABLE 4-1.

Figure 5-12. Sample Failure Status Report'

ence document for failure rates and failure frequency analysis. The failure rates are based upon actual test experience and can be used in making predictions for new systems, as well as for making design and management decisions.

The failure rate compendium also can be used to prepare failure frequency summaries. Failures are caused by design, manufacturing, test, handling, or other factors. Properly organized failure rate compendia can provide considerable insight into the causes of failures.

5-5 TYPICAL OPERATIONAL DATA BANKS

There are four general classes of reliability data centers:

1. The Data Bank/Query System. This system uses a highly structured file to handle logical queries, or perform specific analysis on simple or composite sequences of the data contained in its memory.

2. The Indexing System. This system provides the user with an index (or catalogs) of nonanalyzed reports containing reliability and related information. The user must perform his own analyses and correlations.

3. Structured-input Fixed-query System. This system allows large quantities of raw data to be obtained, in standard formats, from standard procedures. Because of the controlled data collection, a machine usable structure and predetermined analysis are performed to produce a series of periodic statistical summaries. Though the system has the ability to respond to a large variety of queries, formatting and programming requirements often demand that approval be received for nonstandard queries.

4. Analyzed Summaries. This system condenses and analyzes input information to a predetermined extent, and subsequently presents it in periodically updated summaries.

The reports on GIDEP (Part Two, Appendix B) show many methods of using the data banks in that system.

5-6 WARNING

This chapter has elaborated on a comlex, comprehensive system of reports. The reader ought to remember two things:

1. The state of the art in computers is changing so rapidly that minicomputers now can do what the best computers were able to do a decade ago. Nothing in this chapter should be viewed as a restriction on methods of operation. One's own computer department is probably the best source of current information on both hardware and software.

2. The list of "musts" and "shoulds" is long enough so that no program will ever do them all. The important idea is to have a system for keeping track of details and to use all the information available to you to improve the product before it gets to the field. That's what reliability measurement is all about.

REFERENCES

1. AR 750-37, *Maintenance of Supplies and Equipment, Sample Data Collection, The Army Maintenance Management System (TAMMS).*

2. TM 38-750, *The Army Maintenance Management System (TAMMS).*

3. J. Adelsberg, *Reliability and Maintainability Data Source Guide, System Performance Effectiveness Program,* US Naval Applied Science Laboratory, Brooklyn, New York, Lab Project 920-72-1, Progress Report 2, SF013-14-03, Task 1604.

4. W. Grant Ireson, Ed., *Reliability Handbook,* McGraw-Hill Book Company, New York, NY, 1966.

5. MIL-STD-831, *Test Reports, Preparation of.*

6. D. S. Liberman, F. A. Paul, and E. F. Grant, *Failure Reporting and Management Techniques in the Surveyor Program,* NASA SP-6504, 1967.

7. *Guide Manual for Reliability Measurement Program,* NAVWEPS OD29304, US Navy Special Projects Office, 15 May 1965.

CHAPTER 6

ENVIRONMENTAL TESTING

6-1 INTRODUCTION

In environmental testing, an equipment is tested under various environmental conditions of temperature, vibration, radiation, humidity, etc., in order to determine or verify its capability to operate satisfactorily when subjected to stress. Strength, life, and performance tests, as well as other basic test types, may all involve environmental testing. In some hardware development programs, a particular phase of the testing program formally is designated as environmental testing.

The effect of environmental conditions, whether man-made or nqt, on equipment is an important consideration in reliable design. Environmental testing is performed because of the uncertainty of the effects of the environment. The uncertainty of the environment can only be accounted for by conservative design practices and/or by adequate field testing.

In some environmental testing, there is a deliberate attempt to simulate as closely as possible the environmental profile expected during equipment operation. This is occasionally done, for example, in reliability demonstration with samples of prototype hardware. Usually, certain critical features of the total operational environment are simulated at specific severity levels in order to uncover design and material weaknesses and workmanship errors. In still other cases (such as development tests), the operational environment may not be known and test conditions consequently cover a wide range.

All environmental tests have the common objectives of determining the effect of the environmental conditions on an item or verifying that the item is capable of withstanding them. These tests are now employed in essentially all phases of hardware programs from the parts and materials level to large systems. In programs relying primarily on a "build and test" approach, they provide assurance in operational hardware.

An alternative to environmental testing is testing under field conditions. This alternative can provide the desired assurance, but usually costs more (especially in the case of complex, expensive items) and often delays the desired information. In field testing, test conditions generally are not as well controlled as in environmental testing, so that cause and effect relationships may be obscured.

Environmental testing ranges in sophistication from very crude methods, such as using an improvised temperature chamber, to testing in very elaborate facilities permitting many combinations of conditions to be simulated. Tests may be purposely destructive (as in strength and life testing), or nondestructive (such as proof tests and burn-in).

The appropriate test conditions must be carefully selected. Basic factors to consider are:

1. The possible environmental conditions during intended use of the equipment

2. The subset of these that must be used in testing

3. The capability for generating and controlling them.

The environmental factors that can affect the behavior of the item during field operations must be determined. These factors must be simulated to the extent feasible within the constraints of cost, schedule, and testing capability. Not all environmental conditions that affect behavior can be readily simulated, and very rarely can all be generated simultaneously to account for interaction effects. Trade-offs therefore must be made when selecting the test condition.

Ref. 5 is the definitive AMC treatise on the environment. This chapter is a brief summary of a few of the more important points.

6-2 ENVIRONMENTAL FACTORS AND THEIR EFFECTS

Environmental conditions may be, for example, weather and solar radiation, or mechanical shock during transportation and handling, air conditioned rooms for computers, and radio frequency interference. Table 6-1 lists some environmental factors for a typical system.

TABLE 6-1

TYPICAL ENVIRONMENTAL FACTORS'

Acceleration	Meteoroids
Acoustics	Moisture
Aerodynamic heating	Nuclear radiation
Albedo	Pollution, air
Asteroids	Pressure, air
Clouds	Rain
Cosmic radiation	RF interference
Dew	Salt spray
Electric atmosphere	Sand and dust
Explosive atmosphere	Shock, mechanical
Fog	Sleet
Frost	Snow
Fungi	Solar radiation
Gases, dissociated	Temperature
Gases, ionized	Thermal shock
Geomagnetism	Turbulence
Gravity	Vacuum
Hail	Vapor trails
Humidity	Vibration
Ice	Winds and gusts
Insects	Wind shear
Magnetic fields	Zero gravity

The set of environmental conditions that an item encounters during its lifetime is its environmental profile. Therefore, environmental testing must consider all the environments encountered in manufacturing, storage, transportation, and handling, as well as those experienced during operational use.

The environmental conditions are not always known in explicit form. No one knows precisely, for example, the environmental profile that a field artillery rocket will experience throughout its life including all types of environmental factors and their severity levels. It is possible to select representative characteristics, such as averages or maximum levels, of major factors which adequately describe conditions for a test.

Environmental conditions of greatest interest from the reliability viewpoint are those that have detrimental effects on equipment operation. Table 6-2 lists typical detrimental effects of environmental factors. In many cases, effects not detectable when the factors are encountered singly appear when two or more are present simultaneously. For example, some electronic components function properly in either a low temperature or a vibrational environment, but when the environments are combined, component leads break. Some possible combined effects of several environmental factors are illustrated in Table 6-3. Combined environments do not always have adverse effects. For example, low temper-

TABLE 6-2

ENVIRONMENTS AND TYPICAL EFFECTS'

Environment	Effects
Winds, gust and turbulence	Applies overloads to structures causing weakening or collapse; interferes with function such as aircraft control; convectively cools surfaces and components at low velocities and generates heat through friction at high velocities; delivers and deposits foreign materials that interfere with functions.
Precipitation: sleet, snow, rain, hail, dew, frost	Applies overloads to structures causing weakening or collapse; removes heat from structures and items; aids corrosion; causes electrical failures; causes surface deterioration; and damages protective coating.
Sand and dust	Finely finished surfaces are scratched and abraded; friction between surfaces may be increased; lubricants can be contaminated; clogging of orfices, etc.; materials may be worn, cracked, or chipped.
Salt atmosphere and spray	Salt combined with water is a good conductor which can lower insulation resistance; causes galvanic corrosion of metals; chemical corrosion of metals is accelerated.
Humidity	Penetrates porous substances and causes leakage paths between electrical conductors; causes oxidation that leads to corrosion; moisture causes swelling in materials such as gaskets; excessive loss of humidity causes embrittlement and granulation.
Sunshine	Causes colors to fade; affects elasticity of certain rubber compounds and plastics; increases temperatures within enclosures; can cause thermal aging; can cause ozone formation.
High temperature	Parameters of resistance, inductance, capacitance, power factor, dielectric constant, etc., will vary; insulation may soften; moving parts may jam due to expansion; finishes may blister; devices suffer thermal aging; oxidation and other chemical reactions are accelerated; viscosity reduction and evaporation of lubricants are problems; structural overloads may occur due to physical expansions.
Low temperature	Plastics and rubber lose flexibility and become brittle; electrical constants vary; ice formation occurs when moisture is present; lubricants gel and increase viscosity; high heat losses; finishes may crack; structures may be overloaded due to physical contraction.
Thermal shock	Materials may be overstressed instantaneously causing cracks and mechanical failure; electrical properties may be altered permanently.
High pressure	Structures such as containers, tanks, etc. may be overstressed and fractured; seals may leak; mechanical functions may be impaired.

TABLE 6-2 (Cont'd)

ENVIRONMENTS AND TYPICAL EFFECTS'

Environment	Effects
Low pressure (High altitude)	Structures such as containers, tanks, etc. are overstressed and can be exploded or fractured; seals may leak; air bubbles in materials may explode causing damage; internal heating may increase due to lack of cooling medium; insulations may suffer arcing and breakdown, ozone may be formed; outgassing is more likely.
Gases	Corrosion of metals may be accelerated; dielectric strength may be reduced; an explosive environment can be created; heat transfer properties may be altered; oxidation may be accelerated.
Acceleration	Mechanical overloading of structures; items may be deformed or displaced; mechanical functions may be impaired.
Vibration	Mechanical strength may deteriorate due to fatigue or overstress; electrical signals may be mechanically and erroneously modulated; materials and structures may be cracked, displaced, or shaken loose from mounts; mechanical functions may be impaired; finishes may be scoured by other surfaces; wear may be increased.
Shock	Mechanical structures may be overloaded causing weakening or collapse; items may be ripped from their mounts; mechanical functions may be impaired.
Nuclear/cosmic radiation	Causes heating and thermal aging; can alter chemical, physical, and electrical properties of materials; can produce gases and secondary radiation; can cause oxidation and discoloration of surfaces; damages electrical and electronic components, especially semiconductors.
Thermal radiation	Causes heating and possible thermal aging; surface deterioration; structural weakening; oxidation; acceleration of chemical reactions; and alteration of physical and electrical properties.
RFI	Causes spurious and erroneous signals from electrical and electronic equipment and components; may cause complete disruption of normal electrical and electronic equipment such as communication and measuring systems.
Solar radiation	Effects similar to those for sunshine, nuclear/cosmic radiation, and thermal radiation.
Albedo radiation	Albedo radiation is reflected electromagnetic (EM) radiation; amounts depend on the reflective capabilities of illuminated object such as a planet or the moon; effects are the same as for other EM radiation.
Zero gravity	Disrupt gravity-dependent functions; aggravates high-temperature effects.

TABLE 6-2 (Cont'd)

ENVIRONMENTS AND TYPICAL EFFECTS'

Environment	Effects
Magnetic fields	False signals are induced in electrical and electronic equipment; interfered with certain functions; can induce heating; can alter electrical properties.
Insects	Can cause surface damage and chemical reactions; can cause clogging and interference with function; can cause contamination of lubricants and other substances.
Clouds, fog, smog, smoke, haze, etc.	Can interfere with optical and visual measurements; deposition of moisture, precipitation, etc.; enhances contamination; can act as an insulator or attenuator of radiated energy.
Acoustic noise	Vibration applied with sound waves rather than with a mechanical couple; can cause the same damage and results as vibrational environment, i.e., the sound energy excites structures to vibrate.

TABLE 6-3

ILLUSTRATION OF INTERACTING ENVIRONMENTAL EFFECTS'

	Salt Spray	Vibration	Low Temperature	High Temperature
High Temperature	Accelerate Corrosion	Increase Rate of Wear	Mutually Exclusive	
Low Temperature	Decelerate Corrosion	Intensity, Fatigue, Rupture, etc.		
Vibration	No Interaction			
Salt Spray				

ature inhibits the growth of fungi and rain dilutes the corrosive effects of salt spray. A good tabulation of many environmental factors and equipments is presented in Refs. 2 and 5.

A less frequent effect occurs when one environmental condition creates another, e.g., when arcing between switch or relay contacts causes the formation of ozone, thus changing the environment and its effects.

Some conditions cause cumulative nonreversible changes in the equipment; therefore, when considering equipment behavior, the history of environmental exposures must be considered. For example, heating from welding and soldering can cause permanent shifts in device characteristics; mechanical shock can result in permanent dislocation of a lead or a part; and nuclear radiation can cause permanent defects in semiconductor devices. The need for conditioning items prior to environmental testing to simulate the historical effects must be considered. This conditioning is sometimes necessary to assure that the response during the test is representative of that in operational use. Knowing the environmental history is not important when the effects are reversible, but the reversibility of all important responses can be determined only through careful analysis. Ignoring the nonreversible effects that have occurred in previous tests and operations can result in misleading environmental test results. Of course, these effects may be difficult to assess or simulate, but just knowing of their existence can be very valuable for the test designer.

Experience is frequently the most useful guide for selecting the environmental factors and the severity levels and combinations of them to be used in a test. This prior knowledge and experience can help reduce the number of environmental tests needed to ensure the successful operation of the item.

The difficulties associated with common environmental factors – such as temperature, vibration, and thermal shock – nearly always receive attention. Less familiar factors can sometimes be equally or even more important, e.g., hail and insects demand special attention to determine what characteristics and severity levels are required. With hail, for example, if mechanical impact damage is the major effect, the size, shape, velocity, and number per unit area of the simulated hailstones are the important characteristics. On the other hand, the vibration induced by the incident hail may be the most significant factor. Insects can cause both mechanical and chemical damage, and both characteristics must be evaluated.

When there is little available knowledge about the operational environment or its effect on an item, it is often simpler and more economical to test and see what happens than to spend a great deal of time and money on an independent study. This is essentially the "build-and-test" approach, which has limited value for large and expensive items, but, when used with discretion, it can be useful for,new designs or for new applications of old designs.

6-3 SIMULATING ENVIRONMENTAL CONDITIONS

The emphasis on environmental testing has led to the development of some very elaborate test facilities (Ref. 1). Several good surveys of older environmental test capabilities are presented in Refs. 2 and 3. The most frequently used standard for military procurement is MIL-STD-810 (Ref. 4).

It is not always possible to generate complex conditions, even with the most elaborate facilities. Air turbulence, gases, or insect conditions can be difficult to simulate. Many facilities are even limited in their capability to generate complex temperature profiles.

Because of these problems, a great deal of effort has been devoted to developing sophisticated simulation facilities. But there may be other ways to resolve the question. First of all, it is the effect of the environmental conditions that is of interest, not just the conditions themselves. Therefore, substitutes should be considered. For example, pebbles might be used as a substitute for hailstones if mechanical damage from impact is of interest. Or, if vibration induced by hailstones is of interest, then a vibration test already scheduled may be adequate.

Some effects are investigated more easily from a more fundamental level. Also, the environmental conditions sometimes may be separable into fundamental components. For example, a temperature profile may be simulated by high and low temperature levels and thermal shock. In such cases, effects such as nonreversibility , interactions, and aging must be accounted for.

Elaborate environmental test facilities are not always required. Simply heating individual circuit components near a soldering iron may in some cases be more informative than testing the entire circuit or assembly in an oven. And, in the absence of certain capabilities, an improvised test may be better than none at all. For example, mechanical shock may be simulated by dropping the items from a prescribed height.

When facilities do not exist for generating combined environments, combined environment effects must be simulated by using single environments in sequence. If the severity levels of the environments are not set deliberately to damage the equipment, the order of application is determined by whatever is most convenient. When tests cause damage, the order of environments must be considered carefully, First, apply those conditions least likely to damage the specimen. For the mechanical part, humidity and salt spray tests thus logically would be applied before vibration or a mechanical load test. An electronic part usually would be tested by applying vibration before high temperature. Such test sequencing allows the maximum amount of information to be obtained before damage occurs.

Ordering of environments for items composed of both mechanical and electrical parts is not as clear-cut. The same basic criterion still applies, and the ability to repair the item can greatly influence the ordering.

If both single and combined environmental conditions can be generated, it does not necessarily follow that the combined testing is preferable. The final choice of an approach depends on what is to be accomplished with the test and is influenced strongly by factors such as time, cost, skills, and instrumentation.

Combined environment testing has two significant advantages over single environment testing:

1. The ability to investigate the combined effects of multiple conditions; i.e., combined testing, in most cases, more closely approximates the real environment.

2. Several conditions usually can be applied simultaneously in a shorter time than in sequence, due to savings in set-up time. Therefore, combined testing often saves money. The major disadvantage of this approach is that the initial equipment cost for combined testing is higher.

In qualification and acceptance testing, combined environments are preferable. The increased confidence derived from the knowledge that synergistic effects are accounted for frequently permits the use of smaller safety factors.

When testing to relate cause and effect, combined environment testing is used as an extension of single environment testing. During the development phase, initial testing usually is applied to determine the effects of single environments. Combined environments are employed after single environment effects have been determined and combined effects become of interest. Single environment testing also may be preferable in long duration tests due to the impracticality of committing combined environmental test facilities for long periods of time.

REFERENCES

1. "Testing", *Practical Reliability,* Vol. III, NASA CR-1128, Research Triangle Institute, August 1968.

2. W. Grant Ireson, Ed., *Reliability Handbook,* McGraw-Hill Book Company, New York, NY, 1966.

3. T. B. Stephenson, "A Survey of Modern Environmental Testing", Electronic Design News 11, No 14 (November 1966).

4. MIL-STD-810, *Military Standard Environmental Test Methods.*

5. AMCP 706-115 through -119, Engineering Design Handbooks, *Environmental Series, Parts One through Five.*

CHAPTER 7

ACCELERATED TESTING

LIST OF SYMBOLS

A	=	acceleration factor
$\tilde{A}$	=	incremental acceleration factor
Cdf	=	Cumulative distribution function
d	=	factor for Kolmogorov-Smirnov test
g, b	=	state of a system
G	=	equivalent state of a system
L	=	life
MtF	=	Median time to Failure
P_{10}	=	life: 10%will fail before that time
s		implies the word "statistical(ly)", or implies that the technical statistical definition is intended rather than the ordinary dictionary definition
t	=	time
T	=	absolute temperature (also a subscript); total test time
E	=	random variable
X	=	failure rate
τ	=	transformed time
*		implies an unknown parameter
^		implies a point estimate

7-1 INTRODUCTION*

Accelerated testing is a very loosely defined concept; attempts to make it rigorous generally run into difficulties. Loosely speaking, accelerated testing started when someone said, "Let's shoot the juice to it and see what happens." This means, roughly, "Let's treat it worse than we expect it to be treated in ordinary practice and then see what happens." One difficulty is that treating-it-worse does not always mean "shooting the juice to it". For example, electrical contacts behave better as voltage and current are increased (up to a point) and some warmth may improve matters for electronic equipment by helping to reduce the moisture problem.

Accelerated testing in this qualitative sense is something that anyone can do and that everyone does. There is a reasonably firm qualitative foundation for much of it. It is in the quantitative interpretation that troubles begin. These qualitative and quantitative uses of accelerated testing can conveniently be put into four classes:

1. Qualitative – to see what kinds of failures are generated and to decide then if a modification is worthwhile

2. Qualitative – to get a rough, quick idea of whether or not something can stand the gaff

*Large portions of this chapter are adapted from Refs 2 and 3 – they are similar since the accelerated testing portion of Ref. 2 was written by the author of Ref. 3.

3. Qualitative – to see what happens when the user maltreats the device as he probably will

4. Quantitative – to make a prediction about the life under actual operating conditions.

There is little question that accelerated testing is useful for the 3 qualitative measures; so it is mainly the quantitative problem to which this chapter is addressed.

7-2 TRUE ACCELERATION (Ref. 3)

Several definitions[1] for true acceleration appear in the literature, some of which are not very explicit. Most engineers associate true-acceleration with behavior over time. The one given here is chosen for its generality and applicability. Acceleration need not be true to be useful even though untrue acceleration is more difficult to analyze, even qualitatively.

Acceleration is true if and only if the system, under the accelerated conditions, passes reasonably[2] through equivalent[3] states and in the same order it did at the usual conditions. Let $\mathbf{g}(t)$ be the state of the system under usual conditions and let $\mathbf{G}(t)$ be the equivalent state of the system under accelerated damagers ($\mathbf{G}$ is not the state at the accelerated conditions but is the state after being transformed reversibly down to the usual conditions). Then there is true acceleration if and only if:

1. $\mathbf{G}(t) = \mathbf{g}[\tau(t)]$

2. $\tau(t)$ is a monotonically[4] increasing function of the argument

3. $\mathbf{G}(0) = \mathbf{g}(0)$

4. $\tau(0) = 0$[5].

The acceleration factor A is a defined as $A(t) \equiv \tau(t)/t$. An incremental acceleration factor may be defined as $\tilde{A}(t) \equiv d\tau(t)/dt$. True acceleration is illustrated in Figure 7-1(A) for a state vector with a single dimension – resistance ratio of a resistor. It is, of course, nice if $A(t)$ is a constant with respect to time as in the figure and depends in some quite tractable way on the severity level.

Estimates of an acceleration factor will depend on the statistical procedures used to arrive at them.

It is important to recognize the arbitrariness of the definition especially as regards the word, reasonably. In order to have true acceleration, it is only necessary that the things in which we are immediately interested be close enough under the two sets of conditions. To be specific, not all failure modes and mechanisms need be identical down to the last electron orbital.

Generally, the physical condition of the device will be included in the system state either explicitly or implicitly in sufficient detail to permit judgments to be made about its design and construction relative to the failure modes and mechanisms.

The state of a system is not uniquely defined for a physical system; it is defined only for a conceptual model of the system. The

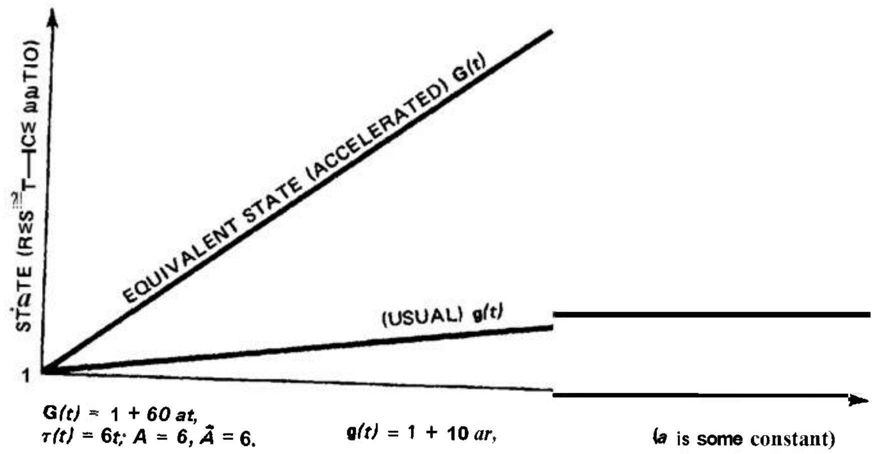
[1] One very poor choice is to assert that acceleration is true if and only if it follows the Arrhenius equation. Another poor choice is to associate it with the constant hazard rate.

[2] The word "reasonably" is necessary because the needs and desires of the situation may be different from time to time, and as engineers, if things are close enough for the purposes at hand, there is no need to worry about the discrepancies as far as these purposes are concerned.

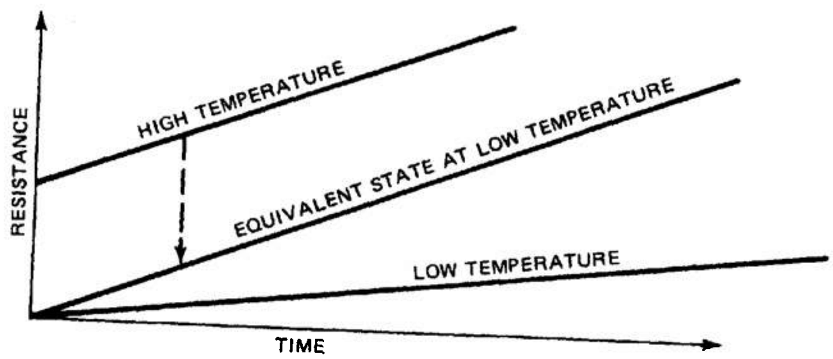
[3] Two states of a system are equivalent if and only if one can be reversibly transformed into the other by changing the environment. For example, a resistor at a higher temperature might never have the same resistance it would at a lower temperature, solely because of its temperature coefficient. This is illustrated in Fig. 7-1(B).

[4] For those who think the term is ambiguous, monotonic is used here in the strict sense, i. e., staying constant is not allowed.

[5] If G and g have a one-to-one correspondence with the argument (the reciprocal function exists), this is a logical consequence of #1 and #3.



(A) Illustration of True Acceleration (see text for notation)



(B) Equivalence of States

Figure 7-1. Accelerated Tests

detailed specification of the system state will vary with our needs and desires and with the required tractability of the resulting equations. The state of a system ordinarily will have several dimensions (components); so it can be classed as a vector. For example, consider a resistor. If we are concerned only about its resistance and nothing else then the state of the system will be given by the resistance of the device (or something equivalent thereto such as a ratio of the resistance to an initial resistance). On the other hand, we may be concerned about the resistance, the temperature coefficient of resistance, the voltage coefficient of resistance, and the chemical composition of the resistive material. Then there will be several dimensions for the system state, and two states will not be the same unless all corresponding dimensions are pair-wise the same.

Lest one be concerned that associating a system state only with a system model rather than with the system itself is too sloppy, an analogy can be made to thermodynamics. There can be many thermodynamic models of a system depending on what is of concern. The entropy is not defined for the system itself but only with regard to a particular thermodynamic model of that system.

In addition to verifying that true acceleration exists, a great deal of effort must be expended in determining the acceleration function. It is virtually always presumed that the acceleration function is a constant.

7-3 FAILURE MODES AND MECHANISMS

Some gross failure modes which can be accelerated for mechanical parts are fatigue, corrosion, creep-rupture, stress corrosion, and various combinations of them. In electronics, one does not ordinarily specify the gross failure modes for acceleration, but, rather, specifies the "stresses" which are being increased. Some of these are temperature, supply voltage, power dissipation, vibration, humidity, and corrosive elements. There is a large body of material in the mechanical and metallurgical fields dealing with those gross failure modes. Since the behavior of electronic components is organized differently, there is no organized body of literature which deals with gross failure modes that cut across all components. A number of information sources on accelerated testing of electronic components are available, many of which are listed in Ref. 3; some of them contain conceptual errors and ought to be read critically. Ref. 3 itself is dated (as are its references) because it was issued in 1968. The failure modes of semiconductors, or at least their relative importance, have changed drastically since then.

7-4 CONSTANT SEVERITY-LEVEL METHOD

This is the traditional type of accelerated test in which the seventy level remains constant throughout the life of the items on test. It is customary to run tests at several severity levels and to plot a curve of some parameter such as failure rate vs severity-level. A sample of several items usually is put on test, and the test stopped when some fraction of the original sample has failed or a specified test time has elapsed. For reliability prediction purposes, the early fraction that fails is most important because only the short-lived items are going to affect the reliability seriously.

7-5 STEP-STRESS AND PROGRESSIVE-STRESS METHODS

The word "stress" is used in the sense of severity level. In this method, the severity applied to a sample of items is increased in steps or increments until some criterion is met for terminating the test. All steps do not have to be the same size, even though this is a common practice.

The term step-stress used in the literature is ambiguous. It is convenient to classify step-stressing into three categories:

1. *Large steps* in which the steps are presumed high enough and long enough so that, for a given step, the damage accumulated at all previous steps is negligible.

2. *Small steps* in which the steps are small enough so that in the analysis one can presume with negligible error that the severity level is steadily increasing. This is then just the progressive-step case; in progressive-stress the stress increases at a constant rate.

3. ***Medium steps*** for which the assumptions for neither small nor large steps are valid. The cumulative damage at previous steps must be taken into account, but the steps are not small enough that the severity level can be considered to be continuously increasing.

In addition, the following terminology is used: large/step-stress, medium/step-stress, and small/step-stress. The size designations are not absolute, but are relative to the kind of analysis that must be performed.

Large/step-stress tests are analyzed as if they were constant-stress tests being run at the severity level of the last step. Parts that are very expensive or otherwise difficult to acquire or test often are treated in this way. Often a sample of only one is used. It is wise to consider the results as "ballpark" figures, since the necessary assumption that the effects of previous steps are negligible is likely to be in error. Preliminary tests often are run in this way and are followed by more comprehensive set of tests later on.

Small/step-stressing is analyzed in the same way as progressive-stressing, and, in fact, by definition, there is really no distinction between them. In many cases, there may be an economic advantage to choosing either very small step increments or a continuously increasing stress. For example, if extremely accurate voltage steps are desired, a stepping switch might be used with a voltage divider; otherwise, a slow motor might be used to turn a multiturn potentiometer.

Less testing time is the major advantage of step-stress tests over constant-stress tests. A direct comparison of the methods requires an assumption of a theory of cumulative damage (see par. 7-6). In the area of metal fatigue, there are many theories of cumulative damage. In electronics, a simple linear model is assumed most,often because of its simplicity and the lack of knowledge of the actual processes.

A linear model of cumulative damage is a gross approximation. In some circumstances, it consistently underestimates and in other circumstances, consistently overestimates the correct results. Regardless of these deficiencies, it offers the advantages of being tractable, easily remembered, and widely used. So use the linear model unless you know of some other which is better. But remember the arbitrariness of any assumption.

An important parameter in step-stress testing is the ratio of severity step size to the time at each level. This controls the rate of increase of the stress severity and is the parameter that is varied when running several tests on a particular population of items.

For some kinds of items, the maximum useful severity level will be exceeded before the device fails in the proper mode. For example, on thermally stressed transistors there are sometimes eutectic points where melting occurs and the transistor essentially ceases to be a transistor. If this happens, the rate of increasing the stress severity must be decreased. The slope of the steps during the course of the tests also can be modified. The severity level limits (i.e., the level where the device ceases to function in its usual manner) are an important limitation to step-stressing. There are other cases where the failure mode changes so drastically at some level, that it is senseless to continue testing above that level.

Another advantage of step-stress testing is the elimination of "switch-on" problems, such as initial transients and failures due to high stress rates. This is because the severity level is low at the beginning and the severity increase is gradual. The severity level need not begin at "zero" (i.e., a level near benign). This can save time and reduce the amount of cumulative damage at severity levels other than the failure level. Some programs concerned with investigating cumulative damage theories may change the seventy level only once during a test. For example, the initial part of one test may be at a high severity level and the remainder at a low severity level; a subsequent test reverses the procedure. Not much work of this sort is done in electronics, but metallic fatigue is a field in which these methods of programming stresses have received considerable attention.

7-6 CUMULATIVE DAMAGE

In order to compare tests (or field experience) run under different severity-level programmings, some model of cumulative damage is necessary. No particular model is required, merely some model. In electronics there are very few theories of cumulative damage, regardless of the part, but in mechanical fatigue, for example, there are many models of cumulative damage. Most often such a model uses constant-"stress" test as its basis. The most common conceptual model, in almost any field, for cumulative damage is the so-called linear model. It has one basic assumption, i.e., the rate of doing damage is $1/MtF$ where MtF is the Median time to *F*ailure[1]. The MtF is for the particular severity level at which damage is accumulating. There are several corollaries[2] to this assumption which often are (but improperly) stated as additional assumptions:

1. The rate of doing damage does not depend on the amount of damage already done.

2. The order in which the severity levels are applied makes no difference.

3. The total damage is the simple sum or integral of the damage done at each stress level.

4. The rates of doing damage are independent of each other for different severity levels.

5. The Median' endurance at constant severity level is unity.

With regard to Corollary No. **5**, the actual endurance is $1 + \varepsilon$, where ε is a random variable; its statistical properties depend on the programming of the severity levels, on the probability distributions of the times to failure at each severity level, and on the percentage chosen in footnote 2. It usually is presumed that the calculated life is the Median (or the percentage in footnote 2).

The use of a cumulative damage model does not necessarily mean that the failure modes/mechanisms were the same at each severity level, although such a case may help the validity of the model. Example No. **32**

---

[1] The Median (i.e., 50th percentile) is the conventional fraction to use. Onc could as easily use some other percentile, e.g., 1% (1% have lives less than the given time). The percentile in the definition and in corollary 5 must agree, of coursc.

'Corollarics 1, 2, 4 are truc becausc the damage ratc depends only on MtF, not on time (for No. 1) nor on s severity lcvel ordcr (for No. 2) nor on the value of MtF for some other severity level (for No. 4). Corollary 3 is true bccause total damagc D is

$$D = \int_0^t \frac{dD}{d\tau}\, d\tau = \sum_i \left(\frac{dD}{dt}\right)_i \int_{\Delta t_i} d\tau$$

$$= \sum_i \left(\frac{dD}{dt}\right)_i \Delta t_i \,.$$

Corollary 5 is truc sincc the median time to failure, at a given severity level is MtF, by hypothesis; the total (median) damagc is damage-rate ($1/MtF$) multiplied by time MtF which is unity.

Example No. 32

Life of a Transistor when the Temperature Fluctuates

Assumptions:

1. The curve of life (appropriate percentile) vs temperature (at constant temperature) is known.
2. The severity level can be completely characterized by a temperature.
3. Linear cumulative damage is appropriate.
4. No new failure modes, which would decrease the life, are introduced by the temperature changes. (In the theoretical development, this is irrelevant since No. 3 determines the method of calculation. But when wondering whether or not No. 3 applies, this is something to consider.)

Let the P_{10} life (10% will fail before that time) be given by the life curve, Fig. 7-2, and the temperature profile be a regularly repeating pattern as shown in Fig. 7-3.

illustrates the linear cumulative damage hypothesis.

Table 7-1 can be developed from the P_{10} life curve and the temperature profile.

The damage rate is the reciprocal of P_{10} life. The fraction of damage has units of 10^{-8} L/hr where L is the presumed equivalent P_{10} life. This fraction is calculated by multiplying the numbers in the 2 preceding columns. It is from the total of fraction-of-damage column that L is calculated, i.e., the total must be unity. From the P_{10} life curve, it can be shown (for what it is worth) that a constant temperature of 162°C would give the same P_{10} life. It is interesting to compare the "% life column" with the "% damage column", e.g., at 350°C, 15% of the life causes 43% of the damage; while at 50°C, 40% of the life causes less than 4% of the damage.

From the remarks earlier in this paragraph it should be remembered that L = 42.9 X

TABLE 7-1

PERCENT DAMAGE VS TEMPERATURE

	Temp °C	P_{10} life 10 hr	Fraction of Life hr/period	Fraction of Life %	Damage rate 10^{-8}/hr	Fraction of Damage 10^{-8} L/hr	Fraction of Damage %
actual	50	5.0	10 + 6	40	0.20	0.08	3.4
	100	1.0	4	10	1.0	0.10	4.3
	150	0.50	8	20	2.0	0.40	17.2
	250	0.20	6	15	5.0	0.75	32.2
	350	0.15	6	15	6.7	1.00	42.9
equivalent	162	L = 0.429	40	100	2.33	2.33	100

Example No. 32 (Cont'd)

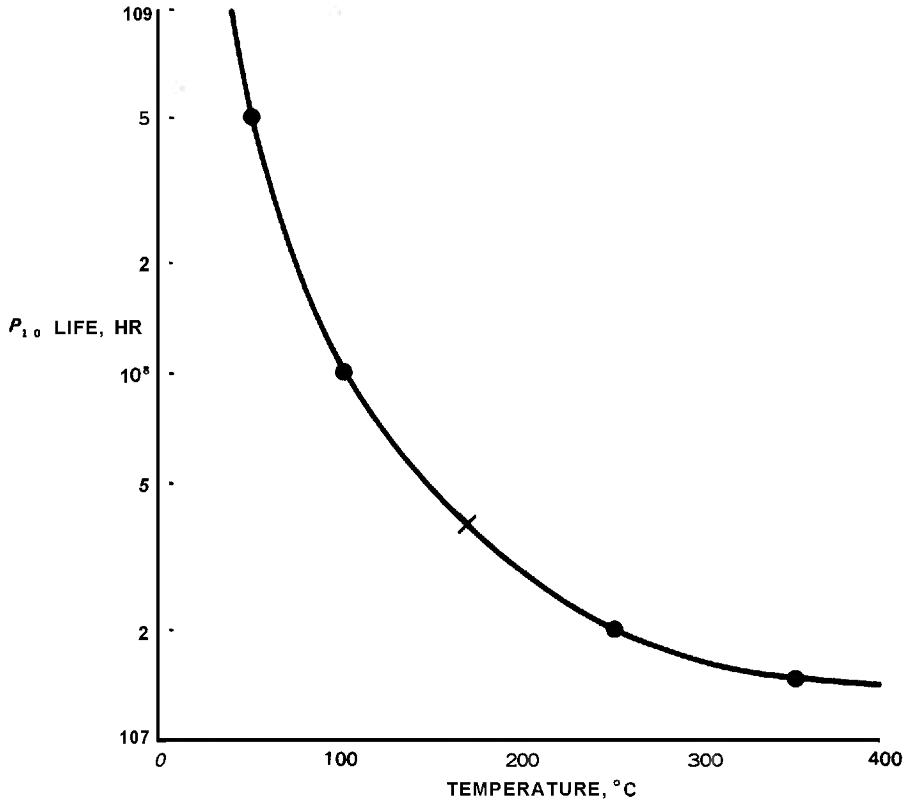


Figure 7-2. Life Curve

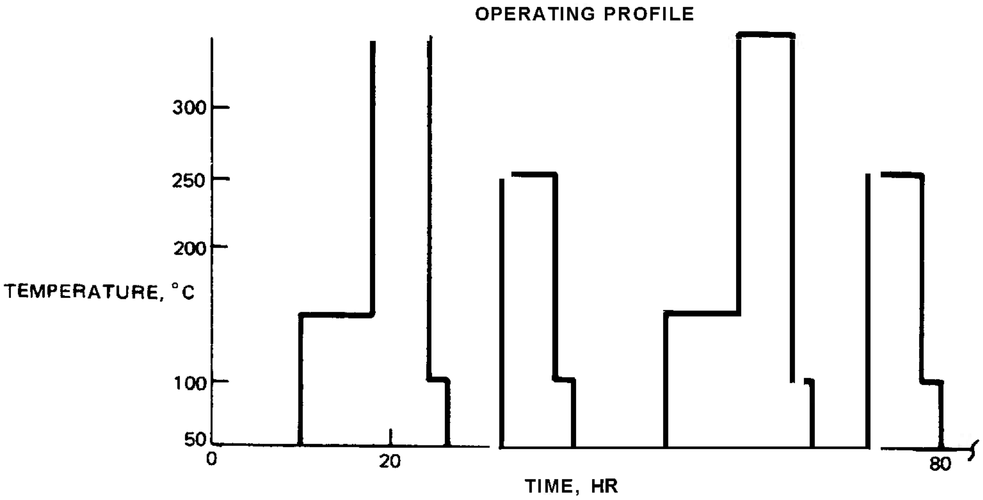


Figure 7-3. Temperature Profile

10^6 hr will not be the actual P_{10} life, but is presumed to be close to it.

7-7 APPLICATIONS

Accelerated life tests can be used in programs operating under tight and critical schedules. If production begins before research and development is completed, some assurance must be obtained quickly that the equipment has an adequate lifetime and that no gross design weaknesses exist. Life tests often take too long to be used under these conditions.

Accelerated testing also is used when repair parts must be manufactured simultaneously with a short run production program. In this case, failure rate data cannot be provided quickly enough by life testing to influence the analysis of the repair complement.

In some cases (explosives, for example), the earliest times to failure in the storage environment and variations in times to earliest failure must be known with high accuracy. A large sample would have to be kept in storage in a usual life test. This would be very expensive and time consuming. Accelerated testing of critical failure modes can be used to determine the range of variability of the time to failure, with useful accuracy, with a smaller sample than that required for usual life testing. For example, solid rocket propellants are subject to catastrophic failure modes that may result in explosions. The remaining life in a stored lot must be estimated periodically, generally every six months or a year. The aged samples must be subjected to accelerated aging in order to determine whether some critical failure mode is about to be triggered.

Accelerated tests can be implemented by speeding up the duty cycle or the environmental level or both. The environments must be cycled from extreme to extreme in order to reproduce in a short time the degradation expected over the period of actual service life. The environmental factor selected for acceleration is determined by the item tested and its failure modes. For example, for many mechanical components, failure is caused by mechanical wear; hence, the acceleration is obtained by increasing the frequency and severity of stress.

The failure data at usual environmental levels and those at accelerated environments must correlate in some way with the stresses actually applied. A precise statistical correlation frequently cannot be obtained because much of the theory of accelerated testing is still very crude. In such cases, accelerated environmental tests may permit a great deal of intuitive information to be developed.

Statistical correlation often can be obtained with accelerated duty cycle testing. The expected number of cycles of actual service in a given time period often can be estimated. Accelerated testing is performed by increasing the number of cycles in a given time period and measuring the mean-cycles-to-failure. The mean cycles between failures (MCBF) at the accelerated duty cycle frequently can be related to the MCBF at normal duty cycles as a function of the ratio of cycles per time period.

7-8 PARAMETRIC MATHEMATICAL MODELS

The ideal accelerated test should include (Ref. 1):

1. An algorithm for converting the reliability data observed at accelerated conditions to reliability data at normal conditions

2. A statistically sound empirical proof of the algorithm

3. A physical model explaining the algorithm.

Unfortunately, most accelerated test techniques do not meet these criteria. They tend to be approximate and require a great deal of engineering judgment in the absence of precise physical models or statistical techniques. Some discussions in the literature are vague and ambiguous; the word "life" may mean the random variable, or may refer to the mean life, or may not mean anything specific. Beware of using acceleration factors from the literature, for specific components. The state of the art in components changes rapidly enough so that failure modes and their acceleration factors can be expected to change in some nonapparent way.

Temperature is the most common method of accelerating a test; usually the scale parameter of the distribution is presumed to follow the Arrhenius law (see Eq. 7-1). Voltage can be increased for some kinds of capacitors, and power dissipation can be increased to shorten the life of many electronic components. Mechanical excitations, such as vibration and shock, are sometimes used.

In the parametric approach, the parameters of the failure distribution are presumed to change in a deterministic fashion with the "stress". The functional relationship of the "deterministic fashion" is presumed known, and the purpose of the test is to evaluate the parameters in that relationship.

The most common situation is the constant failure rate and the Arrhenius temperature law as shown in Eq. 7-1.

$$\lambda_T = \lambda_0 \exp\left(\frac{T^*}{T_0} - \frac{T^*}{T}\right) \qquad (7\text{-}1)$$

where

λ_T = failure rate at T

T = absolute temperature

λ_0, T^* = unknown parameters

kT^* = so called activation energy, where k = Boltzmann constant

T_0 - fixed known temperature

Eq. 7-1 can be put in other algebraically equivalent forms.

Where a failure distribution has 2 parameters, it is most common to assume that one of them is independent of the accelerating "stresses". Other assumptions tend to be intractable, even if more realistic. In the s-normal or lognormal distributions, the median usually is assumed to be a function of the "stresses", and the other parameter to be a constant.

No matter what distribution is assumed, it is essential that the statistical uncertainty in the results be estimated and clearly stated – because this uncertainty is usually so large as to greatly reduce the impact of the nominal conclusions. In fact, one is often tempted to remark, "I could have guessed that close without the tests!". For example, using Eq. 7-1 and some reasonably high temperature tests, the uncertainty in failure rate at operating temperatures might be a factor of 10 or so.

Refs. 3 and 4 give some examples of the application of Eq. 7-1 to real data. The maximum likelihood equations for the solution of the problem, and the computer FORTRAN program to effect the solution, are also given there.

Example Nos. 33 and 34 illustrate the procedure. These two examples show how grossly misleading it can be to give only point estimates.

7-9 NONPARAMETRIC MATHEMATICAL MODELS

Suppose it is known that the time scale factor is k times worse under certain severe operating conditions. Then if accelerated tests are run under those severe conditions, the *Cdf* of the time-to-failure under usual operating conditions can be estimated in a rather short time. It will be as if time were passing k times as fast as usual.

Generally, it is the early failures that are important because they show the early part of the failure distribution. If several items are put on test at the same time, the early part of the *Cdf* can be estimated from the first few failures, regardless of the form of the distribution.

See for example, pars. 2-2 and 2-5.

Example Nos. 35 and *36* illustrate the procedure.

Example No. 33

Capacitor life test, fixed calendar time at each temperature. The table shows the test conditions and the raw failure data.

Temperature, "C	Number of devices	Total-test-time, (10^6 hr)	Number of failures
25	100	0.2688	1
70	100	0.2688	0
125	100	0.2688	4
145	100	0.2688	12

The activation energy was estimated to be 0.10 to 0.62 eV; the failure rate at 25°C was estimated to be 0.04 to 10 per 10^6 hr; the range for each is for a total of 4 standard deviations.

The point estimates are 0.363 eV and 0.633 per 10^6 hr. If only point estimates had been given, they would have been very misleading. In truth, one knows relatively little about the 25 "C failure rate of the capacitors, or how the failure rate changes with temperature.

Example No. 34

Npn planar silicon transistors, life test. There were 2 classes for failure modes: ionic and all-other; the results are shown for ionic only, and for all failures combined. The table shows the test conditions and the raw failure data.

Temperature, °C	Total-test-time, 10^6 hr	Failures all	Failures ionic
175	0.125	0	0
200	1.816	1	0
220	0.125	0	0
240	0.125	0	0
265	0.123	3	3
290	0.104	13	13
320	0.102	17	12
350	0.084	38	6

The results follow; the range is for a total of 4 standard deviations:

1. All failures:
   - a. activation energy — 0.85 to 1.22 eV
   - b. failure rate at 25°C — 0.01 to 13.9 per 10^{12} hr
2. Ionic failures only:
   - a. activation energy — 0.61 to 0.98 eV
   - b. failure rate at 25 "C — 0.47 to 635 per 10^{12} hr.

Example No. 35

A newly designed motor-generator set must be tested. Suppose that under severe conditions of temperature, humidity, and load, that the rate of degradation is 90 times as bad as under ordinary conditions; i. e., the time acceleration factor is 90. Ten items are put on test; the first 2 failures are at 26.7 hr and 43.2 hr, respectively. Estimate the actual *Cdf*.

Procedure	Example
1. Use the Kolmogorov s-confidence limits in par. 2-5. Calculate the times-to-failure under usual conditions.	1. Acceleration factor = 90, $t_1 = 90 \times 26.7$ hr $= 2403$ hr $t_2 = 90 \times 43.2$ hr $= 3888$ hr.
2. Find the 90% s-confidence d from Table 2-12,	2. $d = 0.37$ for 90% s-confidence.
3. Use Eq. 2-48 to calculate the envelope for the *Cdf* where i = failure number and N = sample size.	3. a. For $i = 1$: $F_{Hi} = 1/10 = 0.10$, $F_{Lo} = 0/10 = 0.00$; thus for $0 \leq t \leq t_1$, the upper limit for the *Cdf* is $0.10 + 0.37 = 47\%$, and the lower limit is 0. b. For $i = 2$: $F_{Hi} = 2/10 = 0.20$, $F_{Lo} = 1/10 = 0.10$; thus for $t_1 \leq t \leq t_2$, the upper limit for the *Cdf* is $0.20 + 0.37 = 57\%$, and the lower limit is still 0.

As with all tests involving only a few specimens, the results are discouragingly imprecise.

Example No. 36

Same as Example No. 35, but suppose that the failure rate is constant. Estimate the failure rate and the *Cdf* at the 2 failure times. The test was stopped at the second failure.

Procedure	Example
1. Calculate the total-test-time. State the number of failures.	1. $T = 10 \times 2403 \text{ hr} + 9 \times 3888 \text{ hr}$ $\approx 59 \times 10^3 \text{ hr}$.
2. Estimate the failure rate.	2. $\hat{\lambda} = 2/(59 \times 10^3 \text{ hr}) = 33.9/10^6\text{-hr}$
3. Use Table 2-8 to get symmetrical 2-sided 90% s-confidence limits for λ.	3. The 5% and 95% levels are used (95% − 5% = 90%). There are 2 failures. The factors are 0.18 and 2.4. $\lambda_{lower} = 0.18 \times 33.9/10^6\text{-hr} = 6.1/10^6\text{-hr}$ $\lambda_{upper} = 2.4 \times 33.9/10^6 \text{ hr} = 81/10^6\text{-hr}$.
4. Calculate the *Cdf* limits at 2403 hr and at 3888 hr.	4. At 2403 hr: upper $Cdf = 1 - \exp(-81 \times 10^{-6} \times 2403) \approx 18\%$ lower $Cdf = 1 - \exp(-6.1 \times 10^{-6} \times 2403) \approx 1.4\%$ At 3888 hr: upper $Cdf = 1 - \exp(-81 \times 10^{-6} \times 3888) \approx 27\%$ lower $Cdf = 1 - \exp(-6.1 \times 10^{-6} \times 3888) \approx 2.3\%$.

This is less uncertain than in Example No. 35 (due to the use of parameters in this example) but is still not very good.

REFERENCES

1. W. Yurkowsky, et al., " Accelerated Testing Technology", Vol 11, *Handbook* of *Accelerated Life Testing Methods,* RADC-TR-67-420, Hughes Aircraft Company, November 1967.

2. "Testing", *Practical Reliability,* Vol. 111, NASA CR-1128, Research Triangle Institute, August 1968.

3. R. Evans, *Literature Review Study on Accelerated Testing* of *Electronic Parts* prepared for Jet Propulsion Laboratory, California Institute of Technology under Contract No. 951727, April 1968.

4. R. Evans, "The Analysis of Accelerated Temperature-tests", *Proceedings of 1969 Annual Symposium* on *Reliability,* Jan 1969, pp. 294-302.

CHAPTER 8

NONDESTRUCTIVE EVALUATION

8-1 INTRODUCTION

Nondestructive evaluation (NDE) does not degrade the item to which it is applied; the phrase "nondestructive testing" (NDT) is used to describe similar activities, but is usually considered to be less general than NDE. Techniques such as infrared scanning and X-ray radiography can be employed to:

1. Prevent destruction of items while measuring properties that usually would require destruction if measured by conventional techniques, and/or

2. Permit certain measurements to be made more rapidly and conveniently than conventional techniques allow.

The major application areas of NDE are illustrated in Table 8-1. In research and development, NDE can be used to measure and evaluate special properties of materials. Useful process control information can be generated from NDE which monitors the production process. NDE is valuable in quality control where items and materials can be evaluated without destructive sampling. NDE also can be used to measure wear and deterioration of in-service items.

In NDE, items are observed, measured, exposed to **X** rays, magnetized, vibrated, acoustically excited, heated, etc. No one form of energy nor any one NDE method can answer all NDE requirements. Each technique has its limitations and many of the methods complement one another. On some projects, it may be necessary to develop a special NDE method in parallel with the development of the system to be tested.

Most NDE methods do not measure a parameter or characteristic directly, but measure some more easily observed phenomenon that can be correlated with the desired characteristic. For example, the uniformity of a material can be inferred by observing magnetic flux perturbations or ultrasonic energy reflections. On the other hand, methods like X-ray radiography permit more direct observation. Table 8-2 summarizes typical characteristics of many NDE methods.

8-2 OPTICAL METHODS

Optical techniques use microscopes, magnifying glasses, interferometers, etc. to detect the presence of surface flaws, anomalies, and malfunctions. **A** permanent record of surface conditions and outward appearance can be obtained by photography. This method can provide excellent permanent records, but can only detect and record surface phenomena.

Microscopes provide a maximum magnification on the order of 2000 with field-of-view and depth-of-field decreasing with increasing magnification. Interferometer microscopes offer depth measurements in the low micrometer region. The capabilities of the microscopic approach are greatly expanded by using electron microscopes.

Optical microscopy and photographic techniques can be combined to produce

TABLE 8-1

APPLICATIONS, FUNCTIONS, AND EXAMPLES OF NDE[1]

Areas of Application	Function Performed	Examples
Research and Development	Evaluating materials, components, and parts; comparing and evaluating fabrication and assembly techniques; data acquisition.	Measuring fatigue in metals, detecting cracks in welds, and non-bonds in bonded materials.
Process Control	Measuring process variables and providing control information	Radioisotope thickness gauging.
Quality Control	Detecting and locating anomalies in materials, defective parts, etc.; detecting and locating fabrication and assembly defects; evaluating the production process.	Poor adhesive bonding, cracks in welds, contaminated transistors, non-uniform porosity in metals.
In-service Evaluation	Detecting flaws, defects, wear, and deterioration of items in field use without major disassembly.	Locating corrosion inside gas tanks, detecting moisture in bonded wing structures on aircraft, etc.

photomicrographs – photographs taken through microscopes. Fiber optics technology can be used to observe and record information in otherwise inaccessible areas, such as the inside of fuel tanks or completed wing structures. Wide-angle and long-range photography permit a large amount of information to be recorded. High-speed photography can produce records of the dynamic characteristics of a material or item. Optical equipment is available on an off-the-shelf basis from many sources.

Optical holography is a rapidly expanding field for NDE, providing 3-dimensional imaging (Ref. 8), and offers substantial promise. It is generally quite complex and will require experts in the field.

8-3 RADIOGRAPHY

Radiography is a method for examining the bulk of solid objects. A radiographic system includes three major components: the radiation source, the radiation detector, and the material or item to be inspected. The basic arrangements of these components are illustrated in Fig. 8-1. The arrangement in Fig. 8-1(A) is the more common. Radiation passes through the object onto a film or other radiation detector. The presence of flaws, anomalies, and foreign objects is revealed by the image or detector-output; it is a shadow-casting process. In the other arrangement, the source and the detector are placed on the same side of the material (Fig. 8-1(B)); radiation passes through the detector and strikes the material, causing scatter or secondary emissions which are then detected.

Both nuclear and atomic radiation are used in radiographic NDE. Some important characteristics of this technique are summarized in Table 8-3.

X- and gamma radiation, using sensitive film detectors, are widely used. These methods can detect defects on the order of 1% of the material thickness. Procedures have also been developed which use the Polaroid and Xero-radiographic processes. Color radiography is available; this technique permits hue and saturation to be distinguished, in addition to brightness, so that areas of opacity can be distinguished more easily.

TABLE 8-2

CHARACTERISTICS OF POPULAR NDE METHODS'

Method	Advantages	Disadvantages	Application	Will Detect
Optical	Applicable to almost any item; both dynamic and static measurements; versatile.	Can detect surface phenomena only; sometimes requires a delay (development of photographs).	Surfaces of materials and items; interior of vessels and compartments (fiber optics).	A wide range of surface flaws; visible damage, etc.
Radiography	Can detect hidden and internal defects; both static and dynamic measurements.	Requires expensive and complex equipment; sometimes presents a radiation safety hazard.	All materials not adversely affected by the incident radiation; moving machinery; enclosed objects; internal characteristics.	Flaws in totally enclosed components, machinery, etc.; defects in fast moving equipment.
Thermal	Provides temperature profile of a surface or area during operation; rapid; very sensitive (R).	Sometimes requires application and removal of special materials; permanent records are difficult to obtain (except with IR).	Surfaces of material and items not damaged by application of the coating; very small surfaces such as microcircuits (IR)	Small temperature gradients on surface.
Magnetics	Can locate flaws or defects in assembled equipment; rapid and simple to apply; sensitive.	Applicable to only magnetic materials.	Surface of metals; wires, tubes, etc.	Flaws and anomalies on surface or within magnetic materials.
Liquid Penetrants	Inexpensive and simple.	Requires application and removal of special materials; can detect surface flaws only; a development process precedes output information.	Surfaces of materials and items not damaged by the process; nonporous materials.'	Surface cracks, flaws, and defects.
Ultrasonics	Penetrates deep; can detect flaw with access to only one side of material; sensitive.	Requires a specially skilled operator; permanent records and readout difficulty.	Bonded structures and materials; all solids.	Disbonds, defects, cracks, and flaws.

IR = infrared

*Figure 8-1. **Basic Arrangements of Radiographic Measurement Components'***

Electron radiation is generated by electron accelerating tubes (electron guns) and by X or gamma rays entering a material. Electron beam radiation is used with electron microscopes to map subsurface phenomena. Resolutions in the submicrometer region have been achieved. Electron radiography also makes use of the source-detector-material arrangement shown in Fig. 8-1(B). An advantage of this method is that access is required to only one side of the material. Electron radiation also can be transmitted through the material to measure material density and thickness.

TABLE 8-3

CHARACTERISTICS OF RADIOGRAPHIC NDE METHODS'

Radiation	Source	Material (Applications)	Detector
X Rays	Conventional X-ray equipment	Locating foreign objects, flaws, and anomalies in parts and materials; observing machinery in operation.	X-ray sensitive films; color radiographic films; Polaroid process; Xeroradiography; fluoroscopy.
Electrons (Beta Particles)	Electron guns; secondary electrons emitted when X rays enter sample; radioisotopes.	Density measurements; thickness measurements surface phenomena detection not distinguishable under visible light.	Photographic films; electrical detectors.
Neutrons	Fission reactors and special neutron sources.	Used in lieu of X rays for heavy materials; for use with materials which absorb X rays but not neutrons.	Photographic film sensitive to neutrons; neutron detectors.
Protons	Accelerators	Thickness and density measurements.	Proton detector.
Gamma Rays	Radioisotopes	Used in same manner as X rays.	Photographic films; gamma detectors.
Alpha Particles	Radioisotopes	Thickness measurements of very thin materials.	Alpha detectors with electrical readout.

Neutrons are absorbed differently by different materials (nuclei). These differences sometimes can be exploited for better flaw discrimination. Most heavy materials do not absorb neutrons well, so that thick sections can be investigated with shorter exposure times. Neutrons can be used to examine materials (such as many plastics) which contain hydrogen. A disadvantage is that neutrons are difficult to record on film, and a special process is required for detection.

Alpha and beta particles are both used for thickness measuring. By using a wide range of energies, thickness measurements using beta sources can be made from 15 pin. of aluminum to 0.050 in. of steel. Alpha particles have been used to measure 1% thickness changes in thin foils and paper (Ref. 2).

Radiography is probably the most widely used NDE method. Equipment for these tests is readily available. The topic is thoroughly described in Ref. 3.

8-4 THERMAL METHODS

The flow of heat through a material is altered by discontinuities in the material. These discontinuities produce variations in temperature at the surface of the item. The location and size of an anomaly can be determined by the temperature profile at the surface. Thermal methods are useful for evaluating bonds between two materials.

The heat can be applied artificially or may be generated by the operation of the item. For example, engine cylinders can be uniformly heated by being filled with hot oil. An operating microcircuit produces heat internally. Many methods of detecting and recording the surface temperature gradients are available. Table 8-4 outlines the characteristics of these techniques.

The "frost" test can be used for testing the bond quality of clad nuclear fuel elements. A chemical that has a frosty appearance and a known melting temperature is applied to the element and heat applied. A poor bond causes a change in appearance. The method can be used on other materials.

The temperature profile of a surface can be sensed by a coating of a phosphor suspended in a liquid. The reaction of the phosphor to ultraviolet light changes as a function of temperature.

Tempilstiks® are crayons made of a material that melts at a calibrated temperature. A specimen is marked with the appropriate Tempilstik® before heat is applied. The mark melts at its calibrated temperature. Other similar Tempil products, such as Tempilaq® and temperature sensitive pellets, are available commercially.

Temperature-sensitive paints (Thermocolor, Ref. 3) which change color as a function of temperature also can be used. Some of these paints change color as often as four times at four different temperature levels. The changes are permanent and provide a good permanent record. These paints can be used on almost any surface over a range of 104° to 2912°F and are accurate within ±9 deg F. The paint must dry for 30 min before use and must be removed after use. The surface to which it is applied must be thoroughly clean before application.

Infrared photography and photomicrography can be used to record temperature profiles of surfaces. A newer and more sophisticated method is infrared scanning. Here, the surface of the specimen is scanned by an optical-mechanical system which focuses small portions of the surface onto a IR detector (Ref. 4). This technique can be used to determine the temperature profile of microcircuits. Temperature differences as low as 0.5 deg *C* can be resolved, and components separated by as little as

TABLE 8-4

CHARACTERISTICS OF THERMAL METHODS'

Method	Application	Detector	Detector Applied by	Capabilities	Disadvantaaes
Frost Test	Cladding bond quality of nuclear fuel elements and other bonds.	Acenaphthene or Diphenyl.	Brush or spray.	Can locate flaws with dimensions 0.1 $in.^2$ X 0.2 in.	An important flaw can have low thermal resistance and go undetected.
Temperature Sensitive Phosphor	Evaluating metal-to-metal bonds, fusion bonds, etc.	Zinc-cadmium sulfide phosphor in a plastic suspension viewed under ultraviolet light.	Brush or spray.	Changes emission by 20% when temperature changes 2% over range of 40" to 130°F.	Must be viewed under ultraviolet; emission change is reversible.
Tempilstik®	Almost any surface.	Temperature sensitive crayons having calibrated melting points.	Marking or touching.	Indicates a temperature within tolerance of ± 1% over a range of 113" to 2000" F.	Indicates only one temperature per application; will not work at higher temperatures in a reducing environment.
Temperature Sensitive Paints	Almost any surface, e.g., metals, ceramics, stone, porcelain, plastics, wood, and glass.	Paint changes color with temperature-change; is permanent.	Brush or spray.	104" to 2912°F within ± 9 deg F; can locate flaws on order of 0.001 in.	Must dry 30 min before use.
Infrared	Any surface emitting IR radiation, i.e., heated surface.	Photographic film and IR detector.	No contact; IR applied to detector or films by optical lens system.	Can detect temperature differences as low as 0.5 deg C; has a resolution as small as 0.0014 in.	The more sensitive methods require expensive equipment.
Temperature Probes	Temperature measurements of surfaces and bulk of most materials.	Thermometers, thermistors, thermocouples, and resistance thermometers.	Contact to material or medium to be measured.	Can measure temperatures between approximately −200° to +2000°C within ± 1%.	Has low resolution for temperature profiling; lead wires conduct heat away from surfaces, etc., reducing true temperature.

IR = infrared.

0.0014 in. can be distinguished. The method has been used to determine bond quality in large objects such as solid-fuel rocket motor cases (Ref. 5). A practical application of the IR technique is discussed in Ref. *6*. The big disadvantage is that the emissivity of the surface must be constant.

IR photographic equipment can be obtained from many photographic equipment manufacturers. A sensitive solid-state IR detector also has been developed.

Temperature probes can be used as temperature profile gages. These probes use conventional thermometers, thermistors, resistance thermometers, and thermocouples as the temperature sensing elements. These devices measure temperature accurately, conveniently, and economically; however, large numbers of them are required to profile a surface without reducing resolution. **Also,** the devices themselves, and associated lead wires, etc., tend to lower the temperature to be measured. Therefore, the temperatures recorded with these devices are somewhat lower than these recorded by no-contact techniques such as IR scanning.

8-5 LIQUID PENETRANTS

Liquid penetrants can be used to detect surface flaws in most materials. Surface flaws are open to the surface but are not readily detectable by visual means. Few flaws are revealed by penetrant inspection which could not be seen visually, but penetrants make the defects much easier to locate. Penetrants can be used with metals, glazed ceramics, plastics, and nonporous materials. Specialized penetrants are available for porous materials.

Flaws are located by covering the surface of the material with a liquid having a low surface tension and a low viscosity. The liquid is drawn into the surface defects by capillary action. After the excess penetrant is removed from the surface, a developer is applied which makes the penetrant and the flaw visible.

There are two basic types of penetrants: dye penetrants and fluorescent penetrants. Dye penetrants consist of a dye dissolved in the liquid penetrant. The color of the dye is chosen to give greatest contrast with the developer. One particular dye penetrant provides a red-on-white record of defects which can be removed from the material as a permanent record.

Fluorescent penetrants consist of a fluorescent phosphor dissolved in the liquid penetrant. This type of penetrant works in the same manner as other penetrants. However, flaws must be viewed under near-ultraviolet light *(365* nanometers).

Some precautions must be taken while using these materials:

1. The surface of the specimen must be thoroughly cleaned before the penetrant is applied.

2. Sufficient time must be allowed for the penetrant to penetrate the flaw.

3. The excess penetrant must be removed with care.

4. The developer must be applied in a specified temperature range.

Two special penetrant techniques use radioactive penetrants and filtered particles. The radioactive method uses a radioactive penetrant and detects the amount of this penetrant trapped in defects by either a photographic method or with a radiation detector. This technique is used primarily to determine the porosity in metal alloys. The filtered particle method is used to detect flaws in porous surfaces such as concrete or carbon. The penetrant, in this case, contains suspended particles. The liquid is absorbed by

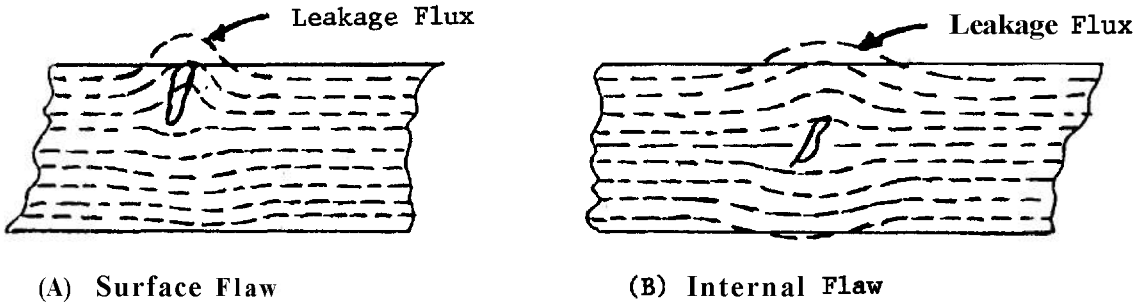


Figure 8-2. Effect of Flow of Magnetic Flux Lines'

the defect, but the particles are larger than the defect and are filtered out and left behind on the surface. These particles then give an indication of a flaw. Fluorescent particles can be used to provide more contrast.

Liquid penetrant inspection is covered in detail in Refs. 3 and 7.

8-6 MAGNETICS

This method is based on the fact that flaws in a magnetic material have magnetic properties different from those of the material itself. Once a magnetic field is induced in a material, any flaw will perturb or distort the field because the flaw has a different magnetic permeability and, thus, a different reluctance, than the material. The flaw is located by measuring these perturbations. Fig. 8-2 shows how flaws affect magnetic flux lines.

A magnetic field can be set up in a magnetic material by:

1. Passing a current through all or a portion of the specimen

2. Passing a current through a coil surrounding or in contact with the specimen

3. Permanent magnets.

The method depends on the type of flux lines desired. Passing a current through a specimen generates circular flux lines around the current path. A coil around a specimen and magnets produce longitudinal magnetization. Both types of flux lines may be needed because the extent to which a flaw perturbs flux lines depends on its orientation with respect to the direction of the lines. For example, a crack perpendicular to flux lines perturbs them whereas a crack parallel to the lines may not. Thus, both flux orientations may be necessary to detect all flaws in a material.

Several methods can be used to detect the perturbations caused by flaws and defects. The simplest is to pass a compass over the magnetized surface. The compass needle will align with the overall field except in the vicinity of a flaw. Although this method is crude and insensitive, the same principle gives good results when extended to distributing iron filings – either dry or in a liquid suspension – over the surface of interest. These filings sometimes are coated with a fluorescent material that produces a more visible pattern. The filings line up with the induced magnetic field, except in the area of flaws or discontinuities in the material. Another detection method is to pass a current-carrying search coil over the surface. When the coil moves through a perturbation, a voltage is generated between the coil and the inspected material.

The magnitude of the voltage gives an indication of the size of the flaw. This method is very useful when the entire object to be inspected (pipes, wire, etc.) can be passed through a coil. A third method takes advantage of the Hall effect. Hall effect probes usually are made of a semiconductor material and are passed over the surface of the magnetized specimen. Variations in the magnetic field due to defects and discontinuities result in a variation in the Hall voltage of the probe.

The sensitivity of this method depends on the strength of the magnetic field. Defects of any consequence can usually be detected down to 0.060 in. below the surface. Defects down to 0.100 in. deep will show under ideal conditions (Ref. 1). A number of other factors such as sharpness, direction, and orientation of the defects also affect the sensitivity of the method.

Eddy current testing is based on the principle that when a coil carrying a high frequency alternating current is brought into the vicinity of an electrical conductor, currents are induced. These currents, in turn, induce a magnetic field about the conductor. The induced currents and the magnetic field are affected by the permeability of the material. Eddy currents can be used to test and measure hardness, alloy content, uniformity of heat treatment, as well as for flaw detection. Flaws perturb and distort the magnetic field produced by the eddy currents.

Two general types of probes are used in eddy current testing. One is an encircling coil which surrounds the specimen and investigates everything within the coil geometry. The other is a point probe which inspects only the area beneath it.

The coil type detector is affected by all the metal enclosed by the coil, so statements about sensitivity are difficult to make, e.g., a long shallow crack may give the same output as a short deep one. At maximum resolution, a defect whose length is comparable to the coil length (roughly 0.1 in. minimum) and whose depth is 5% or more of material wall thickness can be detected. The probes are also sensitive to the displaced volume of metal and are sensitive to defects on the order of the probe diameter. One such probe, the Probolog, developed by Shell Development Company, can detect cracks or seams 0.005 in. deep by 0.5 in. long. The probe also can detect 1% thickness changes in a 0.5 in. length.

8-7 ULTRASONICS

Ultrasonic waves are acoustic waves above the audible range. They are employed in NDE to detect and locate flaws in composite materials and nonbonded areas in bonded materials. The impedance to ultrasonic propagation is different for a flaw or anomaly than for the basic material. Therefore, a portion of the induced ultrasonic energy is reflected by a flaw, just as it is by a boundary of the material. Measurement of the reflected portion or the unreflected portion is the basis for employing ultrasonics in NDE.

There are three methods of ultrasonic testing in general use: pulse echo, transmission, and resonance.

8-7.1 PULSE ECHO METHOD

In the pulse echo method, an applied pulse travels through the material and is reflected from flaws and material boundaries. As the surface of the material is scanned, the appearance of a "defect pulse" locates the surface position of a flaw. The energy of this pulse is related to flaw size, but is usually difficult to correlate precisely. By monitoring the time relationship of the initial pulse, the "defect pulse", and the echo pulse, the defect can be located in depth. Many techniques have been used in the pulse echo method. For example, by introducing the initial pulse at an angle to the material surface, the boundary reflection can be effectively removed in the return. Also, flaws that are not accessible by a simple geometry can be detected by letting the pulse zig-zag from one boundary to another until a flaw is reached. Thus, rather complex geometries can be probed by this method.

8-7.2 TRANSMISSION METHOD

The transmission method is very similar to the pulse echo technique. A pulse is introduced at one boundary of a material and sensed at another. The energy transmitted past the flaw is attenuated due to reflection. The energy received is, therefore, less when a flaw is present than when there is no flaw. This energy decrease indicates the surface position of a flaw, but gives no measure of depth.

8-7.3 RESONANCE METHOD

In the resonance method, a material is excited at its thickness resonant frequency. The material is driven by a transducer which is, in turn, controlled by a variable frequency oscillator. When the resonant frequency is reached, a standing wave will be established between the material faces. As the surface of the material is scanned, any change in resonant frequency not associated with material thickness indicates a flaw. This method is used for thickness measurement as well as for flaw detection. Both longitudinal and transverse ultrasonic waves are used. Special propagation modes called Rayleigh waves and Lamb waves are less frequently used.

Rayleigh waves are surface waves analogous to ripples on water and can be generated by controlling the angle of incidence of the input ultrasonic energy. The Rayleigh wave technique is useful for scanning across the surface of an item for flaws near the surface. A distinct advantage of this method is its usefulness in investigating curved surfaces.

Lamb waves are elastic vibrations analogous to ripples in the whole material. These waves have proven useful in detecting nonbonded areas in laminated structures where vibration in localized areas induced by the Lamb waves can be sensed. The Lamb wave technique is capable of detecting cracks that extend as little as 0.001 in. below the surface of a material.

8-7.4 ADVANTAGES AND DISADVANTAGES

One of the major advantages of ultrasonic test techniques is their ability to penetrate into a material to locate flaws. This depends on available power and sensitivity of the detection equipment; however, the technique has been used to locate flaws as deep as 30 ft down in a metallic bar. It also permits rapid measurements and is economical, relatively sensitive, and reasonably accurate for measuring flaw extent and position. Accessibility to a single surface is adequate for detecting many flaws and anomalies.

The resolution of ultrasonic test methods depends on the frequency of the ultrasonic propagation. The higher the frequency, the smaller the defect that can be resolved. A limiting factor is that absorption of ultrasonic energy increases with increasing frequency. Therefore, a trade-off between frequency and available energy must be made. Generally available equipment permits detection of flaws with dimensions in the 0.001 to 0.005 in. range.

The inconvenience of getting ultrasonic energy into and detecting the energy from a specimen is one of the major disadvantages of ultrasonic methods. Air does not provide the proper impedance match between the transducer and specimen, so that liquid couplants such as oil, water, or glycerine are required. Another major disadvantage is in readout of the information. Display of pulse positions on a CRT often is used. Two-dimensional scanning and imaging is being employed to a limited extent but requires further development to become a practical tool. Major drawbacks of the imaging techniques are distortion and limitations on resolution. Another disadvantage is that highly skilled operators are required. Specimen geometry limits the size, contour, and complexity of shapes that can be tested. Misleading responses may be obtained from usual internal structural characteristics such as large grains or porosity.

REFERENCES

1. "Testing", *Practical Reliability*, Vol. III, NASA CR-1128, Research Triangle Institute, Research Triangle Park, NC, August 1968.
2. W. McGonnagle, *Nondestructive Testing,* McGraw-Hill Book Company, New York, NY, 1961.
3. R. C. McMaster, *Nondestructive Testing Handbook,* Ronald Press, Inc., 1959.
4. M. Walter, J. Roschen, and E. Schlegel, "An Infrared Scanning Technique for the Determination of Temperature Profiles in Microcircuits", IEEE Transactions on Electron Devices, **10** (1963).
5. C. E. Betz, "Nondestructive Testing – Recent Achievements", Industrial Quality Control **23,** No. 1, (July 1966).
6. B. Selikson and J. Dimauro, *Reliability Screening Using Infrared,* RADC-TR-66-360, Rome Air Development Center, October 1966.
7. C. E. Betz, *Principles of Penetrants,* Magnaflux Corporation, Chicago, Il, 1962.
8. J. McDermott, "Finally: A Practical Application for Holography", *Electronic Design,* February 15, 1969.

CHAPTER 9

TEST EQUIPMENT

9-1 INTRODUCTION

The success or failure of a test program often depends on how the test equipment is selected, designed, procured, and tested. The test equipment determines the accuracy of the measurements, the repeatability and usability of the results, and the cost of the test program; and these factors in combination frequently determine whether or not the test program is worthwhile. It is very important that great care be taken in choosing the test equipment. Since test equipment is frequently more complex than the hardware being tested, the design of the special equipment should be given to experienced engineers.

A test equipment program should receive the same level of management consideration as the design and production of the hardware. Often, more management attention is required on the scheduling aspects of the test equipment program than on any other part of the program because of the complex nature of the equipment. On complex programs, a test equipment coordinating committee, which functions from the earliest phase of the project until production is firmly established with proofed and approved test equipment, should be appointed. This committee may be required to supervise the design and development of new and unique test equipments, to delineate design requirements which ensure compatibility of test equipments and hardware, and to establish schedules.

Test equipment includes the equipment providing inputs to the hardware being tested, the measurement equipment detecting the output, and the equipment providing the environment to which the hardware is exposed during the test.

Ref. 2 is a good source of material on this topic.

9-2 COMPARATIVE FEATURES

It is convenient to consider several features of test equipment separately, namely: purpose, type of control, Calibration, and readout.

Test equipment can be classified as either general or special purpose, depending upon whether the equipment is usable on one or more than one type of test article. General purpose test equipment should be used whenever possible, unless some feature of the test program mandates the use of special purpose equipment. Among the factors which may dictate special purpose equipment are:

1. No general purpose equipment is commercially available to make the test.

2. General purpose equipment error is too large and consumes too much of the product tolerance.

3. General purpose equipment set-up time is too long, considering the frequency with which the proposed test will be per-

formed, and the general purpose equipment use factor is too high to permit tying it up in a permanent set-up.

4. Test time with general purpose equipment is excessive, and the frequency of the test performance is high enough to warrant the cost of designing and building special equipment.

Usually, general purpose equipment provides greater flexibility than special purpose equipment, but at a reduced testing efficiency. In large projects the sheer numbers of instruments and test equipments that must be used make it possible to choose special test equipment whenever the economics of a particular test dictate, since the large pool of standard equipment usually available provides all the flexibility required. In small laboratories, problems arise because the supply of standard equipment is limited. It is frequently (though not always wisely) decided to use standard equipment, which is more costly than special test equipment, in order to force the purchase of additional standard equipment.

Test equipment also can be classified according to the way in which it is controlled or programmed, either manually or automatically. As a general rule, initial and maintenance costs of manual control are lower, but it is more expensive to operate because manual control requires constant operator attention. In most repetitive testing, the additional costs of automation can be recovered if the testing continues for a sufficient length of time. Another important saving provided by automatic testing results from the elimination of operator error. Automated test equipment usually provides more repeatable results, permitting easier understanding and diagnosis of test failures or anomalies (which represents additional cost savings), more-uniform testing, and more-consistent quality of product. Therefore, for repetitive testing, the use of automated test equipment should be considered. Substantial savings often can be realized from the automation of cyclic vibration testing.

The form of the output from automated tests should be compatible with the existing computer facilities in the plant to permit rapid data analysis or a Separate minicomputer should be used. Equipment to provide almost any kind of output is commercially available. A printed or visual display can provide an immediate record of results for on-the-spot analysis. This equipment also can show the accept/reject limits simultaneously with.the observed values, and can be programmed to identify an out-of-tolerance condition.

With automatic test equipment, the software can be extremely expensive. Up-to-date knowledge of computer trade-offs is needed, and computer programmers will be a vital part of the project. Flexibility in allowing the program to change with needs should be considered as important as the initial program.

9-3 STANDARDIZATION OF TEST EQUIPMENT

When the product is tested for the same attributes as it passes from vendor to purchaser or user; or when it is tested in several different locations or at multiple field-assembly sites; or when the same tests are used for acceptance testing, field evaluation testing, and repair-depot testing, each at a different location; then, the test results must be compatible. In these situations, the testing and test conditions must be as identical as possible to prevent testing errors and differences from different test locations from affecting the results.

The errors and differences introduced by test equipment at different locations can be important. The test equipment should be identical at all locations. If this is impossible, the differences should be held to an absolute minimum and should be known and evaluated carefully to determine the exact quantitative differences. There are several potential sources of error. The most important are those introduced by the use of supposedly identical or interchangeable equipments which are not really identical or interchangeable. This problem can be minimized by assigning the design of the test equipment to a single group and by instructing the design group to use the same testing techniques and circuits throughout the project. The uniformity of purchased test equipment is more difficult to control, since competitive procurement procedures lead to pressures to accept apparently or allegedly alternate and interchangeable instruments without complete comparative evaluation or analysis. Test personnel must constantly resist this pressure – the world is rarely as advertised and the difficulties can be very subtle and insidious.

9-4 TEST EQUIPMENT ERROR

All measuring equipment has an inherent error. The amount of the error is frequently an important portion of the allowable or desired tolerance on the parameter being tested, so that considerable degradation of the product can result from ignoring test equipment error. Test equipment error exists not only in the reading or sensing portion of the test equipment, but in the portions providing the inputs and environmental conditions as well.

The total error always should be measured for complex special test equipment. The measurement is best made at the interface between the test leads and the article to be tested (where it will include the error of the cabling and connectors), with the equipment and the article both energized and operating. Theoretically-computed errors can be useful during the test equipment design phase, but actual measurements should be made to verify the calculations before the test equipment is released for use. In such calculations, provided they are verified, it is generally permissible to use statistical techniques to combine the individual errors introduced by the many elements of the test equipment.

9-5 TEST EQUIPMENT CALIBRATION

Automation in complex weapon systems has tremendously increased the importance of calibrated test equipment. Because of the increased number and dispersal of the suppliers, manufacturers, and field activities that test all or part of these systems, test engineers must give consistent and compatible results.

Another major source of test errors is a poor calibration system which is not rigorously traceable to the standards of measurement held by the National Bureau of Standards (NBS).

Test equipment incompatibilities can result in major catastrophic failures.

Calibration is defined as the comparison of the indication of a measuring device with a known standard, the known standard itself being calibrated against more accurate standards in a series of rigorously controlled echelons up to a national standard held by the NBS. All test and measuring equipment used in high-reliability projects, including those used by R&D personnel, must be Calibrated against laboratory standards at specified intervals. These intervals should be set after thorough analysis of drift rates or susceptibility of instruments to inaccuracies from handling.

A system of mandatory recall usually is required to ensure satisfactory operation of a calibration program, since most test operators are slow to release instruments for calibration. Mandatory recall means that the calibration laboratory is directed by project management to remove physically an instrument, when due, from the test floor to the laboratory for calibration. To make such an operation practical, a loan pool of calibrated instruments must be available for replacement of instruments removed.

Most large industrial firms and Government activities have their own calibration laboratories. Project test personnel must ensure that the laboratory is advised early in the project of any new (or more difficult) measurement requirements, so that additional standards can be procured and calibration procedures prepared and proofed out in time to support these new measurements.

In smaller companies and activities, however, calibration laboratories are not usually economically feasible. These activities must use the services of commercial calibration laboratories. They should do so with great care, however, because many commercial companies do not have adequate calibration facilities. When no suitable commercial calibration laboratory is available in a local area, arrangements should be made with the nearest Department of Defense industrial facility to perform instrument (or standards) calibration.

The *Standards Laboratory Information Manual,* published by the Department of the Navy, Bureau of Naval Weapons, Pomona, California, presents a complete treatise on calibration and standards laboratories (Ref. 1). This manual includes recommended recalibration frequencies for a large list of commercial measuring equipment. Three basic kinds of calibration are discussed: (1) calibration of individual instruments, such as meters, gages, and power supplies; (2) calibration of systems of complex test or environmental equipment; and **(3)** calibration of standards.

Individual instruments usually are transported to a calibration laboratory for calibration at intervals ranging from one to three months. **A** sticker must be applied to the instrument to indicate the date calibrated and the date next due. Floor inspection personnel have the responsibility of ensuring that all instruments in use in their activity area are in calibration. If a loan pool of instruments is available, the loan instruments should be maintained in an in-calibration condition and should be cycled into the laboratory for recalibration at the same intervals as instruments in use on the floor. When individual instruments are part of a complex test set-up which is in constant use, it is feasible to calibrate only the scales on the instruments which are actually used in the set-up, provided that the instruments are marked as being usable only on those particular scales.

With large fixed instruments, or with delicate moving coil instruments used for highly accurate measurements, it may be necessary for calibration personnel to carry fixed standards from the calibration laboratories to the instrument into the laboratory. If this is necessary, optimum working conditions should be provided in the work area to ensure maximum accuracy.

Calibration of complex measuring and environmental systems usually is performed in-place, although some consoles may be transported to the calibration laboratory. The calibration should be performed at the test leads or at the point of application of the environment so that the errors introduced by cabling and switching or by the input fixtures of environmental equipment will be detected in the calibration process. For control purposes, doors, panels, and removable instruments and equipments should be sealed and break-of-inspection procedures established to ensure that any changes or tampering will be detected and recalibration performed. The recalibration interval of complex equipment must be established by careful analysis of drift data, but it usually can be set initially to correspond with the shortest recalibration interval of any standard equipment installed in the system. Detailed procedures, including before and after data sheets, are essential to calibrate such equipment properly. Audit responsibilities should be assigned to an unbiased inspection group to ensure that the procedures are meticulously followed.

Standards can be calibrated in two ways: (1) by cross-checking of like standards held by a laboratory, and (2) by comparison of the standard with a standard of higher accuracy traceable to a comparison with a standard held by the National Bureau of Standards. Cross-checking is relatively easy and inexpensive, and its use reduces the frequency with which the standards must be referred to a higher-accuracy standard. Certification of standards is a common practice. Inspection departments should be given the responsibility of auditing all the standards in the chain from the local-level standard back to NBS to ensure that floor measurements are within the specified accuracy.

Detailed calibration procedures for both commercial standard measuring equipment and special test equipment must be prepared and used. Calibration usually is performed by highly skilled technicians who may believe that they can dispense with written instructions because they know all the steps. The errors and mistakes which result from this overconfidence are frequently catastrophic, particularly since other project personnel assume that the calibrations have been properly performed.

Auditing responsibilities should be assigned to a "disinterested" group to ensure that detailed procedures are prepared and followed by calibration laboratory personnel. These procedures should be prepared in great detail and with extensive use of hookup diagrams. They should include data sheets that require recording both the as-received and as-adjusted readings, and the instrument errors. The as-received readings are essential to permit studies of drift rates for the purpose of adjusting recalibration periods to suit individual usage conditions and specific drift factors. Calibration procedures prepared by instrument manufacturers are not always detailed enough and must frequently be rewritten.

9-6 TEST EQUIPMENT RUGGEDIZATION

Test equipment must be ruggedized to minimize the loss of accuracy in laboratory or field use. Measurements made with inadequately ruggedized test equipment may introduce errors in the test data. In addition, the inadequate test equipment may be incapable of performing continuously and consistently, thus destroying the effectiveness of the test program. Therefore, it is important that considerable attention is paid to ruggedizing the test equipment. When test equipment is ruggedized at a point in the program after some tests have been run, great care must be taken that the new equipment does not introduce undetected changes into the system. A convenient check on this problem can be made by making compatibility runs and testing the same hardware on both the old and the new test equipment and carefully comparing the results before the new test equipment is released for production use.

To minimize the ruggedizing cost, the reliability group should review the initial R&D test equipment designs to ensure that as much MIL-Spec quality material as possible is specified and that the original drawings are as detailed and complete as possible. Although this approach will increase the cost of the R&D test equipment, it will result in considerable net savings to the project by reducing the amount of redesign and rework effort for ruggedization and by improving the overall quality of both the R&D and the production testing efforts.

9-7 TEST FACILITIES

One of the more difficult decisions facing test management personnel is the make-or-buy decision for the many kinds of testing which comprise the reliability test program. The straightforward economic factors involved are not difficult to assess by common methods of cost accounting and analysis, and further discussion is not warranted here. However, some of the many intangibles which are more difficult to analyze frequently override or modify the economic considerations.

Good management practice is to perform as much testing in-house as possible. This ensures the maximum utilization of capital investments and of the work force. Schedules are more easily monitored and a far greater flexibility can be attained. Flexibility is very important for the R&D test program, in which test forecasts can change. Control of test conditions, discipline of test operations, and liaison between the laboratory and other organizational elements are more easily established and maintained with in-house testing. In-house testing eliminates the necessity for training outside personnel in company techniques, standards, methods, and systems of technical and financial management. Also, the company capability and capacity for performing tests generally are improved as the in-house test load broadens and increases.

Balancing factors must be considered, however, which frequently result in a decision to buy test services. Time may be an overriding consideration. The in-house laboratory or test facility may not have the technical capability, personnel skills, or available capacity to perform all of a required test, and the time to obtain them may be longer than the project schedules will permit. Or, even if there is time, the required test may be very special in nature with little likelihood of being repeated, so that it is not economically feasible to establish an in-house capability.

Sometimes, a customer may decide that the testing should be performed in an outside laboratory to provide an unbiased check on project performance, or he may have his own laboratories and decide to perform the testing there.

Particularly in R&D testing, there may be very sharp peaks and valleys in the scheduled or actual testing load, and it may be economical for the company to buy commercial laboratory time to carry the peaks. In a multidivision company, sister divisions may have very low utilization factors and corporate management may dictate the interdivision transfer of a fixed percentage of testing. There is distinct merit in maintaining a group of competent outside laboratories who are familiar with the company and are available on short notice to take care of unplanned emergency requirements.

It is generally prudent not to schedule the in-house laboratories to capacity; there must be a reserve of unplanned laboratory time available for emergency tests. Qualification tests are not intimately connected with the production process and should also be performed in-house, particularly the ambient tests. Requalification tests can be split; those tests which require environments not available in-house should be performed outside.

The reliability department should monitor each vendor's laboratory to provide close liaison and direction and to monitor and audit the vendor's performance. All of the company's in-house disciplines – including calibration of all test equipment, preparation, release, and change control of detailed procedures and data sheets, operator certification, and independent inspection coverage –

should be required of the vendor and enforced. It may be desirable to have the vendor's procedures reviewed, approved, released, **and** controlled in-house.

REFERENCES

1. *Standards Laboratory Information Manual,* Department of the Navy, Bureau of Naval Weapons, Pomona, CA.
2. "Testing", *Practical Reliability,* Vol. 111, NASA CR-1128, Research Triangle Institute, August **1968.**

CHAPTER 10

RELIABILITY GROWTH

L, U = subscripts which imply Upper and Lower uncertainty

$m(t)$ = mean number of failures in the interval $(0, T]$

n_i = number of failures during t_i (par. 10-2.2)

n_T = number of failures in the interval $(0, T]$

t = time

t_i = time at which failure i occurred; test time of unit **i** (par. 10-2.2)

T = time at which prediction is made

$u(t)$ = measure of uncertainty in $\ln \lambda(t)$

β = constant in the equation for $m(t)$

$\lambda(t)$ = failure occurrence rate; peril rate; intensity of Poisson process

λ_i = failure rate of unit i

10-1 INTRODUCTION

Contrary to conventional wisdom, a product when it comes from design engineering, does not have an "inherent" reliability which is then degraded by the manufacturing department. Rather, the product contains weaknesses which are removed by an experience-fix sequence. Experience includes design-development testing, production screening, field use, etc. – anyway that experience with the product is obtained. Fixing involves changes in design details, production methods, inspections, etc.

Not all fixes cure the ills they were supposed to cure, and sometimes fixes introduce new ills that weren't there before. This is one reason that the Army is properly skeptical about claims that product ills have been cured, when there is no physical testing to support that assertion. It is also reason to be skeptical of reliability growth models that can't decrease.

The repeated process of experience-fix is called Reliability Growth. Often if continues throughout the life-cycle of a product. Since reliability growth ought to be present in all programs, it is desirable to have a method of predicting what the reliability will eventually grow to. The prediction method uses a mathematical model, i.e., a set of equations whose parameters can be estimated from historical data on this product or similar ones. There are many such models, only one of which is treated in detail in this chapter.

10-2 DUANE MODEL

The Duane model (Refs. 1 and 2) and modifications thereof are based on the idea that the more resources which are devoted to finding and fixing what is wrong, the better the reliability will be. The model is usually expressed as "The 'log cumulative MTBF' is proportional to the 'log cumulative test-time', if there has been no major redesign." This statement is neither clear, nor complete, in several ways:

1. The concept of MTBF is being used

very loosely. Usually (and improperly) the inverse of failure rate is meant.

2. It is not clear whether the failure rate of the item is constant during the test (Poisson process) or whether the failure rate changes during the test (nonhomogeneous Poisson process—see Ref. **3,** pp. 125-126).

3. The rate at which resources are devoted to the project must stay reasonably constant, as a function of test-time. This requirement eliminates major redesign as well as major shifts in effort.

4. Different (and possibly incorrect) statements of the exact situation reflect the ambiguity in how time is accumulated and how the nonhomogeneous Poisson process is precisely defined.

Only one of the possible Duane models is analyzed; another is merely referenced. The basic assumption of constant effort (resource rate) is probably the limiting factor in a satisfactory prediction; it is difficult to demonstrate, and a poor-fit of the data to the model is sometimes used as evidence that the effort is not constant. The arithmetic in the first model is somewhat easier; so perhaps it deserves to be tried first.

10-2.1 DUANE MODEL NO. 1

The following assumptions apply:

1. The equipment failures are described by a nonhomogeneous Poisson process (the failure rate is a prescribed function of time).

2. Time is the cumulative test-time of all units on test; e.g., if there are **3** units on test, time cumulates at **3** times the clock-calendar rate.

3. The cumulative number of failures is directly proportional to $(\text{Time})^\beta$, where β is a positive constant. (If $\beta < 1$, the reliability is growing; if $\beta = 1$, the reliability stays the same; if $\beta > 1$, the reliability is getting worse.) This is equivalent to saying that log failure rate is a linear function of log time.

Occasionally, this model is referred to as "Weibull" because of the similarity of this nonhomogeneous Poisson process failure-occurrence rate to the Weibull failure rate; but the basic models for the two processes are quite different.

Notation follows:

t = time

T = time at which a prediction is made (can be t_n)

t_i = time at which failure i occurred, a random variable

n_T = number of failures in the interval $(0, T]$, a random variable unless $T = t_{n_T}$.

$m(t)$ = mean number of failures in the interval $(0, t]$

$\lambda(t)$ = failure-occurrence rate (intensity of Poisson process), $\lambda(t) \equiv dm(t)/dt$; sometimes called "peril rate"

β = constant in the equation for $m(t)$

$u(t)$ = a measure of the uncertainty in $\ln \lambda(t)$

implies an estimate or prediction, See Appendix B for details

L, U = subscripts which imply upper and **lower** uncertainty limits on $\lambda(t)$

From Appendix B, which contains the mathematical details, the solution for the parameters of the model is:

$$\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n} \ln \frac{T}{t_i}\right)^{-1} \tag{10-1}$$

$$\hat{m}(t) = n_T(t/T)^{\hat{\beta}} \tag{10-2}$$

$$\hat{\lambda}(t) = \hat{\lambda}(T)(t/T)^{\hat{\beta}-1} \tag{10-3a}$$

$$\hat{\lambda}(T) \equiv \hat{\beta} n_T/T \tag{10-3b}$$

$$\hat{u}(t) = \hat{u}(T)[\{1 + [1 + \hat{\beta} \ln (t/T)]^2\}/2]^{1/2}$$

$$= \hat{u}(T)[[1 + \{1 + \ln [m(t)/n_T]\}^2]/2]^{1/2} \tag{10-4a}$$

$$\hat{u}(T) \equiv (2/n_T)^{1/2} \tag{10-4b}$$

$$\hat{\lambda}_L(t) = \hat{\lambda}(t) \exp [- 2\hat{u}(t)] \tag{10-5a}$$

$$\hat{\lambda}_U(t) = \hat{\lambda}(t) \exp [+ 2\hat{u}(t)] \tag{10-5b}$$

The uncertainty limits on λ ($\hat{\lambda}_L$ and $\hat{\lambda}_U$) are approximate and are chosen at twice the estimated asymptotic standard deviation of $\ell n\ \hat{\lambda}$. They have no exact probabilistic interpretation, but serve adequately for engineers and managers who want to know whether the estimate is within 10% or a factor of 10. Indeed, the factor of 2 on $\hat{u}$ could as easily have been **3.**

For $\hat{u}(T)$ small, i.e., $\hat{u}(T) \ll 1$, the relative uncertainty in $\hat{\lambda}$ is, from **Eqs.** 10-4 and 10-5; $\pm$ **2** $\hat{u}_T = \sqrt{8/n_T}$; thus to know $\hat{\lambda}_T$ within $\pm$ 10% requires about 800 failures (n_T = 800). For most situations this is an outlandish number, which shows that rarely will we know $\hat{\lambda}$ to within $\pm$ 10%.

Statisticians are divided on the worth of running goodness-of-fit tests. The reasons for not paying much attention to them are that their interpretation is narrowly statistical, rather than engineering, and that the result tends to be very sensitive to the size of the sample (number of failure-occurrences). For any physical situation, it is usually possible to take so few data that the model being considered is rarely rejected, or to take so many data that the model is virtually always rejected. Usually, the uncertainty in the projected value of $\hat{\lambda}(t)$ will be disheartening enough.

Example No. **37** illustrates the procedure.

10-2.2 DUANE MODEL NO. 2

The following assumptions apply:

1. There are 2 kinds of time:

   a. The cumulative test-time of all units on test, up to the time of the last fix.

   b. The test time of a unit, as measured from the time of its most recent fix.

2. The failure rate of a unit is constant between fixes.

3. The log failure-rate of a unit when it is fixed is a linear function of log cumulative-test-time (see Assumption 1a).

Notation follows:

t_i = test time of unit i (see Assumption 1b)

T_i = cumulative test-time of all units on test, up to the time of the last fix for unit i (see Assumption 1a). t_i and T_i are independent.

n_i = number of failures during t_i

λ_i = failure rate of unit i

From Assumption **3,** it follows that

$$\ln \lambda_i = h_0 + (\beta - 1) \ln T_i \tag{10-6}$$

Eq. 6 has the formalism of Ref. 4, where the $\beta - 1$ and $\ell n T_i$ of Eq. 10-6 correspond to the E and x_i of Eq. 12 in Ref. 4, respectively. The solution for λ_i and the uncertainties

Example No. 37

Several pieces of the same equipment are being tested in conjunction with a reliability-growth program. After 1200 total hours of testing time, there have been 8 failures. The individual failure times (not shown here) were used to calculate $\hat{\beta} = 0.6$ from Eq. 10-1. What is the current failure rate, and what is it estimated to be if there are 1800 more hours of testing? (Since $\hat{\beta} < 1$, the failure rate is decreasing.)

Procedure	Example
1. State the data (and calculated results) which are given.	1. $T = 1200$ hr $n_T = 8$ $\hat{\beta} = 0.6$.
2. Find $\hat{\lambda}(T)$ and $\hat{u}(T)$. Use Eqs. 10-3b and 10-4b.	2. $\hat{\lambda}(T) = (0.6)(8)/1200\text{-hr}$ $= 4.0/1000\text{-hr}$ $\hat{u}(T) = (2/8)^{1/2} = 0.5$.
3. Find $\hat{\lambda}_L(T)$ and $\hat{\lambda}_U(T)$. Use Eq. 10-5.	3. $2\hat{u}(T) = 2(0.5) = 1.0$ $\hat{\lambda}_L(T) = (4.0/1000\text{-hr})e^{-1.0}$ $= 1.5/1000\text{-hr}$ $\hat{\lambda}_U(T) = (4.0/1000\text{-hr})e^{+1.0}$ $= 10.9/1000\text{-hr}$.
4. Find t after 1800 more hours.	4. $t = 1200$ hr + 1800 hr = 3000 hr.
5. Find $\hat{m}(t)$, $\hat{\lambda}(t)$, $\hat{u}(t)$. Use Eqs. 10-2, 10-3a, 10-4a.	5. $t/T = 3000\text{ hr}/1200\text{ hr} = 2.5$ $\hat{m}(t) = 8(2.5)^{0.6} = 13.86$. $\hat{\lambda}(t) = (4.0/1000\text{-hr})(2.5)^{0.6-1}$ $= 2.77/1000\text{-hr}$ $\hat{u}(t) = 0.5\{[1 + (1 + 0.6 \ln 2.5)^2]/2\}^{1/2}$ $= 0.652$.
6. Find $\hat{\lambda}_L(t)$, $\hat{\lambda}_U(t)$. Use Eq. 10-5. Calculate the ratio of $\hat{\lambda}_U(t)/\hat{\lambda}_L(t)$.	6. $2\hat{u}(t) = 2 \times 0.652 = 1.30$ $\hat{\lambda}_L(t) = (2.77/1000\text{-hr})e^{-1.30}$ $= 0.75/1000\text{-hr}$ $\hat{\lambda}_U(t) = (2.77/1000\text{-hr})e^{+1.30}$ $= 10.2/1000\text{-hr}$.

Example No. 37 (Cont'd)

7, Compare results of steps 3 and 6.	7, $\hat{\lambda}_U(t)/\hat{\lambda}_L(t) = 10.2/0.75 = 13.6$; $\hat{\lambda}_U(T)$ and $\hat{\lambda}_U(t)$ are about the same; $\hat{\lambda}_L(t)$ is about one-half of $\hat{\lambda}_L(T)$, for this problem, So the extra 1800 hr of testing will achieve at most a factor of 2 improvement in failure rate, but might do no good at all,

is more complicated in this case than in par. 10-2.1. and probably isn't worth the extra trouble, since the uncertainties are going to be large in either case. The computer program in Ref. 4 would have to be modified to account for the new variables.

10-3 OTHER MODELS

A large number of mathematical models are possible for reliability growth. Many of them can fit any particular set of historical data reasonably well. The real question, however, is how well they allow you to estimate current reliability and to predict its future value.

Refs. 5 and 6 each compare several models with simulated data. Ref. 5 is limited to binomial data (success-failure). Ref. 6 includes the Duane model in its comparisons. Ref. 7 analyzes the Duane model. Unfortunately, very few references show how to estimate the uncertainty in reliability-growth predictions. No statistical point estimate is worth very much unless it is accompanied by some measure of its uncertainty.

One big difficulty with choosing a model is to be aware of all the answers that are part of the model, rather than part of the data. As an absurd extreme, if the model is chosen so that there's a lower limit on failure rate, one ought not to be surprised at the statistical predictions from the model and data that there is a lower limit on failure rate.

A reliability statistician can help you create a growth model that you are willing to live with. The cost of analyzing the model will generally be negligible compared to the cost of the physical tests. In the absence of such help, one can plot $-\log\hat{R}$ against "some measure of program resources consumed so far" where $\hat{R}$ is the estimate of reliability from the most recent test. If there are a great many data, the form of the curve will be obvious. If there are very few data, the scatter on the graph will emphasize your irreducible uncertainty.

REFERENCES

1. J. T. Duane, Technical Information Series Report 62MD300, General Electric Co., DCM&G Dept., Erie, PA; 1962.

2. J. T. Duane, "Learning Curve Approach to Reliability Monitoring", *ZEEE Trans. Aerospace,* Vol. *2,* 1964.

3. E. Parzen, *Stochastic Processes,* Holden-Day, Inc., San Francisco, 1965.

4. R. A. Evans, "The Analysis of Accelerated Temperature-tests", *Proc. 1969 Annual Symposium on Reliability,* pp. 294-302. Available from the IEEE.

5. R. C. Dahiya, A. J. Gross, "Adaptive Exponential Smoothing Models for Reliability Estimation", *IEEE Trans. Reliability,* Vol. R-23, December 1974, pp. 332-334.

6. T. Jayachandran, L. R. Moore "A Comparison of Reliability Growth Models", *IEEE Trans. Reliability,* Vol. R-25, April 1976.

7. L. H. Crow, "On Tracking Reliability Growth", *Proc 1975 Annual Reliability and Maintainability Symposium,* pp. 438-443. Available from the IEEE.

APPENDIX A

ENVIRONMENTAL SPECIFICATIONS

A-1 INTRODUCTION

Material from Refs. 1 and 2 is presented which permits the engineer to:

1. Keep track of environmental specifications for both military and nonmilitary parts

2. Find the environmental requirements of **a** Military Specification

3. Compare various requirements at a glance.

A-2 A QUICK GUIDE TO ENVIRONMENTAL SPECIFICATIONS (REF. 1)

Trying to keep track of the various environmental requirements of Military Specifications can be a nightmare, particularly when the engineer is working on several different types of equipment simultaneously. To alleviate this problem, the important data from the common specifications (listed in Table **A-1**) have been tabulated and grouped in Table **A-2.**

TABLE A-1

LIST OF SPECIFICATIONS

(The list shows the Standard and revision considered in Ref. 1)

MIL-STD-810	Environmental Test Methods (Equipment)
STANAG 3518	NATO: Environmental Test Methods for Aircraft Equipment and Associated Ground Equipment
MIL-E-4158 (USAF)	Electronic Equipment Ground; General Requirements for
MIL-E-5272 (ASG)	Environmental Testing, Aeronautical and Associated Equipment; General Specification for
MIL-E-5400	Electronic Equipment, Aircraft; General Specification for
MIL-T-5422 (ASG)	Testing, Environmental, Aircraft, Electronic Equipment
MIL-E-16400 (NAVY)	Electronic Equipment, Naval Ship and Shore; General Specification for
MIL-T-21200	Test Equipment for use with Electronic and Fire Control Systems; General Specification for
MIL-STD-202	Test Methods for Electronic and Electrical Component Parts

TABLE A-2

ENVIRONMENTAL SPECIFICATIONS SUMMARY'

VIBRATION

Specification	Method Para.	Curve/Detail	G	5 10 50 90 200 500 1k 2k 3k Hz
MIL-STD-810 B	514-1,5,6	V A,B Y W AA,AB,AQ,C,M D,Z E F	1.3 2 2.5 4 5 10 15 20	5 – 500 sinusoidal
	514-2,3	.36" / .06" K / N J / P H / Q G / R I / S T U	G 2 5 10 '5 20 30 50	5 – 2k sinusoidal
	514.4	G: AE 5.3, AF 7.4, AG 9.3, AH 11.9, AJ 16.9 G: AK 20.7, AL 23.9, AM 29.3, AN 37.8, AP 46.3		50 – 2k random
STANAG 3518	15k		21.5	10 – 500 sinusoidal
MIL-E-4158 C	3.2.12 3.2.28.2			20 – 55 as specified
MIL-E-5272 C	4.7 Proc. VI discont Proc. I,IX,X,XI use XII	Proc. IV,V VIII VII XII B III, XIII XII A II XIV	 15 10 2 2 10 20 20	5 – 50, 90 – 200 circular 100 – 500 5 – 500 2k
MIL-E-5400 J MIL-T-5422 E	3.1.35 3.2.2 1.5 4.2	Parts Curve II III I,IV	.06" 2 5 10	55 5 – 500
MIL-E-16400 F	3.1.1.8.1 4.5.14			5 – 33 (MIL-STD-167, Type I)
MIL-T-21200 G	3.1.35 3.2.16.4	Parts Class 2,3 Class 1-11 I-III I-I,IV	.06" .06" 2 5 10	10 – 55 5 – 500 500

TABLE A-2 (Cont'd)

ENVIRONMENTAL SPECIFICATIONS SUMMARY'

VIBRATION (5, 10, 50, 90, 200, 500, 1k, 2k, 3k Hz)

Specification	Method	Condition	G	Vibration (Hz)
MIL-STD-202	201A		.06"	10 – 55
	204A	A	10	10 – 500 sinusoidal
		C	10	10 – 2k sinusoidal
		B	15	
		D	20	
	214	Let. / I	II	50 – 2k random
		A / 5.2	5.9	
		B / 7.3	8.3	
		C / 9.0	10.2	
		D / 11.6	13.2	
		E / 16.4	18.7	
		F / 20	22.8	
		G / 23.1	26.4	
		H / 28.4	32.3	
		J / 36.6	41.7	
		K / 44.8	51.1	

TEMPERATURE (-80, -40, 0, 40, 80, 120, 160, 200, 300, 400, 500°C)

Specification	Meth./Para.	Detail	Temperature
MIL-STD-810 B	501 Hi	Proc. I	71
		II	68
	502 Lo	Proc. I	as spec.
	503 Shock	Proc. I	-54 – 71
STANAG 3518	15b Hi	f	55
		e	70
		d	85
		c	100
		b	125
		a	155
	15c Lo	d	-10
		c	-40
		b	-54
		a	-62
	15o Shock		-54 – 71
MIL-E-4158 C	3.2.28.1.1	Indoor	0 – 52
		Out, Moder.	-40 – 52
		Out, Cold	-54
		Out, Desert	71
MIL-E-5272 C	4.1 Hi	Proc. I+II	71
	4.2 Lo	Proc. I	-54
		II	-62
	3.1.2 Shock	Proc. I, 1A	85
MIL-E-5400 J MIL-T-5422 E	3.1.1	Class 2	-54 – 55
			-54 – 71
		3	-54 – 95
		4	-54 – 125
MIL-E-16400 F	1.3 1.B.1 1.5.8	Class 4	0 – 50
		3	-40 – 50
		2	-28 – 65
		1	-54 – 65
MIL-STD-202	102A	Cond. B	55 – 65
		A,D	-55 – 85
		C	-65 – 125
	107 B	A	5[illegible] – 85
		B	125
		F	150
		C	-65 – 200
		D	
		E	350

TABLE A-2 (Cont'd)

ENVIRONMENTAL SPECIFICATIONS SUMMARY'

HUMIDITY

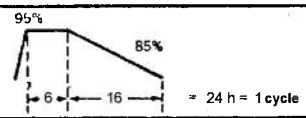
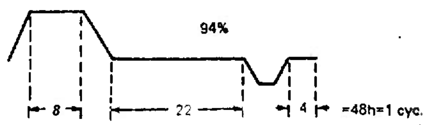
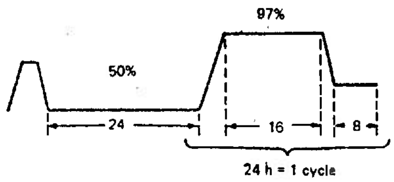
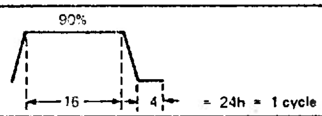
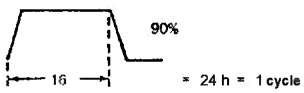
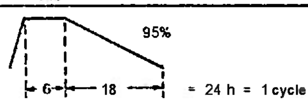
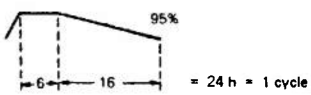
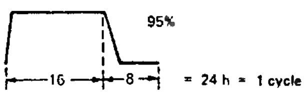
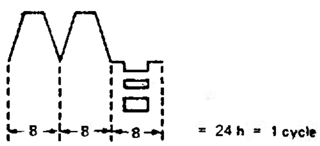
Specification	Meth./Para.	Details	°C	0 12 24 3C 48 hours	No. Cyc.	Total Hours
MIL-STD-810 8	507	Proc. I	71° 28'	 95% 85% 6 16 = 24 h = 1 cycle	10	240
		Proc. II	65' 30' 20°	 94% 8 22 4 =48h=1 cyc.	5	240
		Proc. III		as in Proc II plus 480 h at 30°C		720
		Proc. IV	60" 45° 30" 25°	 97% 50% 24 16 8 24 h = 1 cycle	5	144
		Proc. V	40.5° 21'	 90% 16 4 = 24h = 1 cycle	20	480
STANAG 3518	15e		60" 25'	 90% 16 = 24 h = 1 cycle	5	120
MIL E-4158 C	3.2.28.1		37.8	100%		—
MIL-E-5272 C	4.4	Proc. II, I	71" 28'	 95% 6 18 = 24 h = 1 cycle	10	240
		Proc. III	43'	95%		360
MIL-E-5400 J	3.2.21.4			100%		—
MIL-T-5422 E	4.4		50° 40"	 95% 6 16 = 24 h = 1 cycle	10	240
MIL-E-16400 F	4.5.9		60' 30°	 95% 16 8 = 24 h = 1 cycle	5	120
MIL-T-21200 C	3.2.16.2			100%		—
MIL-STD-202	103 B Steady State		40°	90%	B A C D	96 240 504 1344
	106 B		65' 25° -10°	 8 8 8 = 24 h = 1 cycle	10	240

TABLE A-2 (Cont'd)

ENVIRONMENTAL SPECIFICATIONS SUMMARY'

SHOCK

Specification	Met/Para	Details	Handling	Design	Crash	Hi Impact	Pulse	Notes
MIL-STD 810 B*	516	Pr. II V VI	Drop Bench Rail Imp.					
		Air Ground I Air Ground		15g 11ms 40g 18ms 20g 11ms 40g 18ms			half-s. half-s. sawt. sawt.	
		Air Ground III Air Ground			30g 11ms 75g 11ms 40g 11ms 75g 11ms		half-s. half-s. sawt. sawt.	
		Air Ground IV Air Ground				100g 6ms 100g 6ms 100g 6ms 100g 11ms	half-s. half-s. sawt. sawt.	
STANAG 3518	15j	3a 3b		as spec.	as spc.			
MIL-E-4158 C	3.2.23.2	as specified						
MIL-E-5272C	4.15	Pr. II-V I-IV III		15g 11ms	30g 11ms 50g 8.5	as spec.	half-s.	MIL-S-4456 JAN-S-44 MIL-S-901
MIL-E-5400-J MIL-T-5422 E	3.2.21.6 4.3	}		15g 11ms	30g 11ms			
MIL-E-16400E	3.1.1.8.1 4.5.14.1	}				1,3,5 ft. 400 lb.		MIL-S-901, Gr.A
MIL-T-21200 G	3.2.16.5 4.3.2.1		Transient Drop	15g 11ms	30g 11ms			

SHOCK

Specification	Method	Condition	Handling	Design	Hi Impact	Pulse	Notes
MIL-STD-202	203 A		Drop				
	202 B <4 lbs	as specified					
	205 C <300 lbs	A B C		15g 11 ms 30g 11 ms 50g 11ms			
	213	A B C D E F G H I preferred (D–I)		50g 11ms 75g 6ms 100g 6mr 500g 1ms 1000g .5ms 1500g .5ms 50g 10ms 75g 6ms 100g 6ms		half-s. half-s. half-s. half-s. half-s. half-s. sawt. sawt. sawt.	for semiconductor
	2b7 <300 lbs				1, 3, 5 ft 400 lbs		MIL-S-901, Class HI

TABLE A-2 (Cont'd)

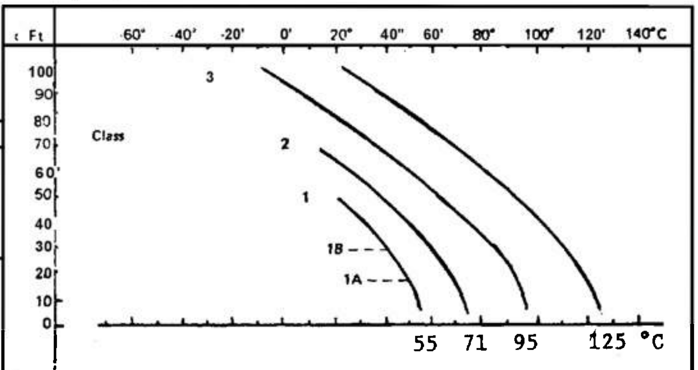
ENVIRONMENTAL SPECIFICATIONS SUMMARY'

TEMPERATURE/ALTITUDE

Specification

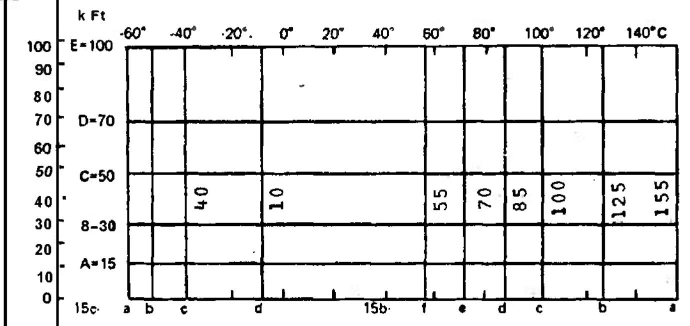
MIL-STD-810 B Meth. 504
Curve A : Cont. Oper.
MIL-STD-810 B Kcth. 504
Curve A . Cont. Oper.

Figure	Class	0 Ft	Alt.	Temp
504-1	1B	55°C	15k	49°C
1	1A	55°	30	40°
1	1	55°	50	19"
-2	2	71°	70	11°
-3	3	95°	100	. 9'
-4	4	125°	100	23'

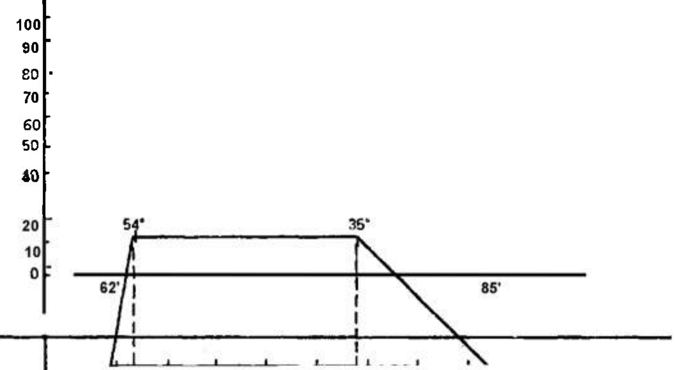


STANAG 3518 15 1.2a

Specify Altitude with a temperature of 15b and 15c.



Proc.	0 Ft	50 k Ft
I	-62° to 85°C	-54° to 35°C
II	as specified	



Class	0 Ft	Alt.	Temp.
1	55°C	50k	20°C
2	71°	70k	10°
3	95°	100k	-9°
4	125°	100k	23°

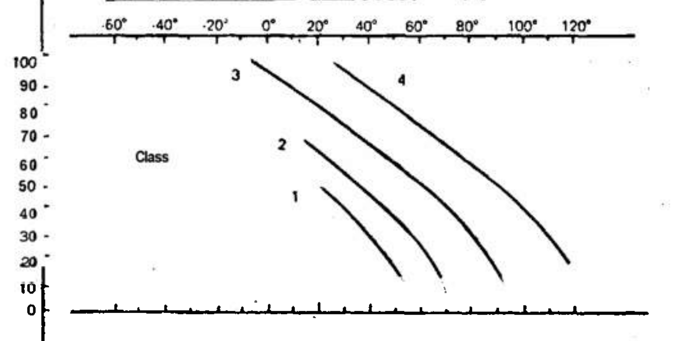


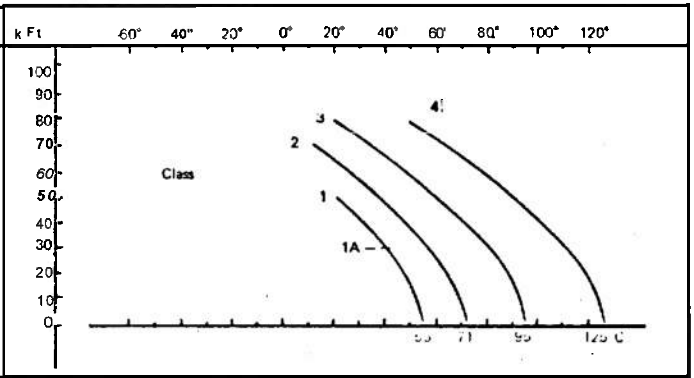
TABLE A-2 (Cont'd)

ENVIRONMENTAL SPECIFICATIONS SUMMARY'

TEMPERATURE/ALTITUDE

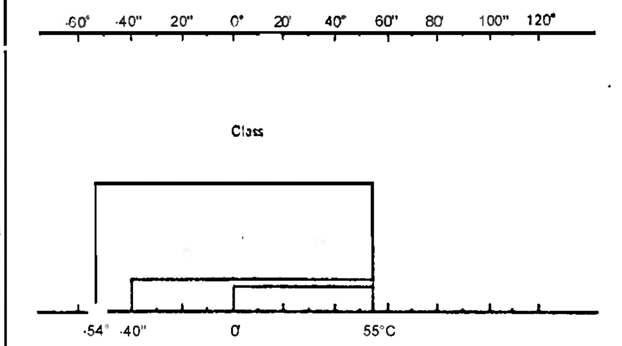
Specification	
MIL-T-5422 E 4.1 Fig. 1 Curve A: Cont. oper.	

Class	0 Ft	Alt.	Temp.
1A	55°C	30k	40°C
1	55°	50k	20°
2	71°	70k	. 11°
3	95°	80k	20°
4	125°	80k	50°



MIL-T-21200 G 3.2.16.1

Class	Altitude	Temperature
3	0 to 10k	0° to 55°C
2	0 to 10k	-40° to 55°
1	0 to 50k	-54° to 55°



Specification	Meth /Para	Detail	20 40 60 80 100 150 656 k Ft
MIL-STD 810 B	500	Proc. I	50 as spec
		II	-- at 54°C
STANAG 3518	15a		15
MIL-E 4158 C	3 2 2 8 1 4		10
MIL-E 5272 C	4.5	Proc. I-VIA	1
		VIB	10
Proc. III, IV are now Proc. II of Temp/Altitude		II-VIC	50
		V-VID	70
		VIE	80
		VIF	100 at -54°C
MIL-E-5400 J	1.2 3.2.21.2	Class 1A	30
		1	50
		2	70
		3,4	100
MIL-T-5422 E	3.1.1	Class 1A	30
		1	50
		2	70
		3,4	80
MIL-T-21200 C			
MIL-STD-202	105C	Cond F	15
		A	30
		B	50
		C	70
		D	100
		E	150
		G	656

Table A-2 is arranged according to the type of test – altitude, temperature, temperature/altitude, vibration, shock, and humidity. Along with quickly spotting a specific requirement for a specification, various specifications for similar tests can be compared. If complete details are needed, then consult the applicable document. Table A-2 makes this even easier by giving the specific paragraph in the specification that has the needed information.

A-3 ENVIRONMENTAL CODE: A SHORT-CUT TO SPECIFICATIONS (REF. 2)

Military Specifications describe, often at great length, the environmental parameters of military electronic parts and equipments. These detailed requirements are needed to ensure accuracy, but they are difficult to record and to compare. A short, simple method of specifying the exact parameters is presented.

The Environmental Code (EC) consists of a 10-digit number. The position of each digit defines a specific parameter, while the value of the digit represents a particular level of that parameter.

For example, the first digit always represents altitude, the second represents high operating ambient temperature, and so forth. If the first digit is a 1, then this specifies 15,000 ft altitude, whereas a 6 represents 100,000 ft. In this way, 10 different environmental parameters are represented in the code (see Fig. A-1).

The Environmental Code Table (Table A-3) lists the various levels for each parameter along with the corresponding code number. In addition, the test methods for parts are listed as per MIL-STD-202. These methods are not the only ones to be used since different parts and equipments use different Military Specifications, some of which are shown in Table A-4. In general, an item to be coded should be tested according to its relevant specification. Since it is the document most commonly used in evaluating electronic parts, MIL-STD-202 is the basis for EC.

Occasionally the levels in Military Specifications are slightly different from the levels in EC. In that case, the deviation is indicated by an asterisk next to the appropriate code number. For example, MIL-E-16400F specifies the temperature + 50°C, while MIL-E-4158C specifies + 52°C. Both of these temperatures would be represented by a 1* in the second position of the EC, indicating that the temperature is close to, but not exactly, + 55°C.

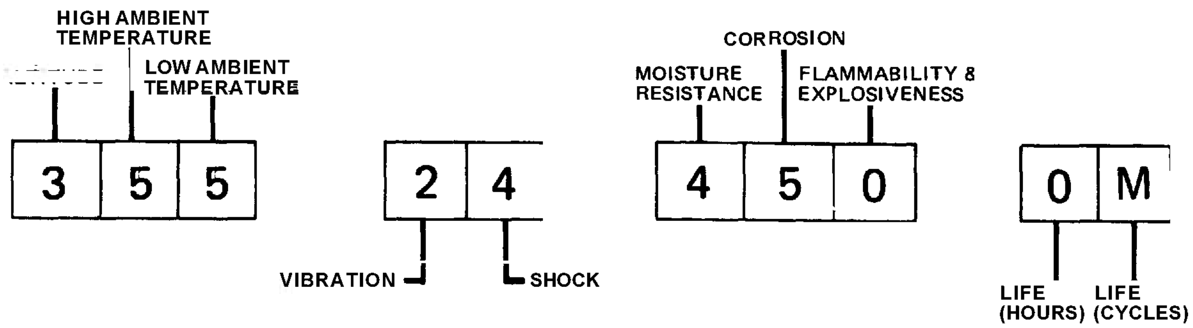


Figure A-1. The Environmental Code (EC)[2]

TABLE A-3

ENVIRONMENTAL CODE TABLE[2]

for Military Electronic Parts and Equipments

ALTITUDE

	kFt	km	in. of Hg	Method 105
0				
1	15	4.3	17.3	F
2	30	9.1	8.9	A
3	50	15.2	3.4	B
4	70	21.3	1.3	C
5	80	24.4	.82	
6	103	30.5	.32	D
7	150	45.7	.043	
8	656	200	9×10^{-8}	E G
9				

HI TEMPERATURE

	°C	Method 102	Method 107
0			
1	+55		
2	+71		
3	+85	A,D	A
4	+105		
5	+125	C	B
6	+150		F
7	+200		C
8	+350		D
9	+500		E

LO TEMPERATURE

	°C	Method 102	Method 107
0			
1	-10		
2	-25		
3	-40		
4	-55	A,D	A
5	-65	C	B-F
6	-75		
7			
8			
9			

VIBRATION

	g	Hertz	Method
0			
1			
2		10-55	201
2	40	10-500	204A
3	20	10-500	
4			
5	10	10-2k	204C
6	15	10-2k	2048
7	20	10-2k	2040
8	30	10-2k	
9			

SHOCK

		Method	
0			
1	Drop	203	
2			
3	15g, 11ms	205A	
4	30g, 11ms	2050	213H
5	50g, 11ms	205C	
7			
8			

MOISTURE RESISTANCE

	Hrs	Method	
0			
1	96	103B	Humidity
2	240	103A	(40°C)
3			
4	240	106	Moist. R
5			(25° to 60°)
6			
7			
8			
9			

CORROSION

	Salt	Sand	Fungus
0			
1	101B, 48h		
2	101A, 96h		
3		110A	
4	101B + 110A		
5	101A + 110A		
6			Fungus
7	101B + 110A + Fungus		
8	101A + 110A + Fungus		
9			

FLAMM. & EXPL.

	Method Flamm.	Method Expl.
0		
1	111	
2		109
3	111 +	109
4		
5		
6		
7		
8		
9		

LIFE (Hours) — Method 108 (at high ambient temperature and full load)

0			
A	96	J	50k
B	250	K	
C	500		
D	1k		
E			
F	2k		
G	3k		
H	5k		
I	10k		

LIFE (Cycles) — Method 206

0			
A	500	K	300k
B	2k	L	500k
C	5k	M	1M
D	10k	N	2M
E	15k		
F	25k		
G	50k		
H	100k		
J	200k		

Levels

Designate the required level by the appropriate digit.

Test Methods

Test Methods specified are for parts as in MIL-STD-202. For Equipments, use the applicable methods of MIL-STD-810.

TABLE A 4

USE OF MILITARY SPECIFICATIONS~

Military Specifications

Name	Contents	Used For
MIL-STD-202	Test methods for electronic and electrical component parts	Parts in general
MIL-STD-750	Test methods for semiconductor devices	Semiconductors
MIL-STD-883	Test methods and procedures for microelectronics	Microelectronics
MIL-STD-810	Environmental test methods	Equipments

By coding the Military Specifications themselves, the coded parts and equipments can be compared to the specifications. The Military Specifications need only be coded once, since they apply as long as the specification remains unchanged. Some sample codings for parts and equipment specifications are shown in Table A-5.

It **is** easy to code Military Specification parts. But what about non-military (Non-MS) parts? They can also be coded if acceptable test evidence exists. By comparing the EC's for these parts with the code of an equipment, parts evaluation can be simplified.

For example, if the equipment EC is

3 14-24-400-DO

and the **EC's** of two parts are

Resistor (MIL-R-11F)	324-75-400-DO
Resistor (Non-MS)	334-53-400-FO

then a comparison will show that the digits of the resistors are equal to or higher than those of the equipment. Hence, both resistors meet the environmental requirements of the equipment.

If however, some of the digits of a part are lower than the corresponding digits in the equipment code, it is immediately known that the part is not suitable for the equipment.

This environmental code was developed in 1963 and is used by the Canadian Military Electronics Standards Agency (CAMESA) for its evaluation procedures of nonstandard parts. CAMESA is the Canadian military agency for electronic parts.

The original code contained other parammeters such as reliability, sunshine, rain; and radiation; but these have been deleted for the sake of simplicity.

In general, EC provides an accurate, easy way to define the environmental characteristics of electronic parts and equipment. It greatly simplifies parts evaluation, especially when large numbers are involved. It is a simple and effective information retrieval system, and can be used either manually or with computers.

TABLE A S

SAMPLE ENVIRONMENTAL CODES[2]

Parts Specs		Environmental Code
MIL-R-11F, Resistor		3 2 4 – 75 – 4 0 0 – D 0
MIL-R-93D, Resistor		6 5 5 – 75 – 4 6 0 – D*0
MIL-STD-446A,	Group I	1*1 4 – 14 – 4 5 3 – J 0
	Group II	4 3 5 – 5 5 – 4 5 3 – J 0
	Group III	1*5 5 – 5 5 – 4 5 3 – J 0
	Group IV	6 5 5 – 5 5 – 4 5 3 – F 0
	Group V	6 7 5 – 65 – 4 5 3 – J*O
	Group VI	7 7 5 – 65 – 4 5 3 – F 0
	Group VII	7 8 5 – 75 – 4 5 3 – F 0
	Group VIII	7 9 5 – 8*5 – 4 5 3 – 1 0

Equipment Specs		Environmental Code
MIL-E-5400J,	Class I	3 1 4 – 2*4* – 1 7 2 – 00
	Class IA	2 1 4 – 2*4* – 1 7 2 – 00
	Class 2	4 2 4 – 2*4* – 1 7 2 – 00
	Class 3	6 4*4 – 2*4* – 1 7 2 – 00
	Class 4	6 5 4 – 2*4* – 1 7 2 – 00
MIL-E-16400F,	Class 1	0 2*4 – 1 7 – 4*1*0 – 0 0
	Class 2	0 2*2 – 1 7 – 4*1*0 – 0 0
	Class 3	0 1*3 – 1 7 – 4*1*0 – 0 0
	Class 4	0 1*1 – 17 – 4*1*0 – 0 0

REFERENCES

1. R. Wernick, "A Quick Guide to Environmental Specification", *The Electronic Engineer,* March 1969.
2. R. Wernick, "Environmental Code: A Shortcut to Specifications", *The Electronic Engineer,* September 1969.

APPENDIX B

ESTIMATES FOR RELIABILITY-GROWTH, DUANE MODEL

LIST OF SYMBOLS

AVar, ACov = asymptotic variance and covariance, respectively

c = $\beta \ln T$

C = $\beta \ln t$

EL_{ij} = s-expected value of second derivative of L with respect to the variables i and j; α corresponds to either i or j = 1 and β corresponds to either i or j = **2**

L, U = subscripts which imply Lower and Upper bounds, respectively

L^*, L = Likelihood, $-\ln$ Likelihood, respectively

$m(t)$ = mean number of failures in $(0, T]$

n = number of failures in $(0, T]$

s = implies the word "statistical(ly)" or implies that the technical statistical definition is intended rather than the ordinary dictionary definition

t_i = time of failure i

T = fixed time of test

$u(t)$ = measure of uncertainty in $\ln \lambda(t)$

a = positive constant

β = positive constant

$\lambda(t)$ = $dm(t)/dt$, peril rate

Π_i, Σ_i = product and sum, respectively, over i from **1** to n

The Duane Model-1 is a particular nonhomogeneous Poisson process wherein

$$m(t) = \alpha t^{\beta} \tag{B-1}$$

Notation follows:

$m(t)$ = mean number of failures in $(0, t]$

α, β = positive constants

$\lambda(t)$ = $dm(t)/dt = \alpha\beta t^{\beta-1}$; peril rate

T = fixed time of test

n = number of failures in $(0, T]$,a random variable

t_i = time of failure i, a random variable

Π_i, Σ_i = product and sum, respectively, over i from 1 to n

L^*, L = Likelihood, $-\ln$ Likelihood, respectively

The statistical procedure used here to estimate α, β, and their uncertainties is Maximum Likelihood. The Likelihood is the probability and/or probability density that the actual results would be found.

Parzen (Ref. 1, p. 143) has shown that:

1. The conditional *Cdf* of the failure time is

$$Cdf\{t \mid n; T\} = m(t)/m(T), \quad \text{for } 0 \le t \le T;$$

the condition is that the number of failures n is known.

2. The failure times behave as if they were s-independent.

The $Pr\{n;T\}$ is the same for a nonhomogeneous process as it is for a homogeneous one, i.e.,

$$Pr\{n; T\} = \exp[-m(T)][m(T)]^n/n! \quad \text{(B-2)}$$

The Likelihood is

$$L^* = [\Pi_i pdf\{t_i \mid n; T\}] Pr\{n; T\} . \quad \text{(B-3)}$$

It is tedious but straightforward to show that (the random variables, t_i and n, are the constants; the 2 parameters a and β are the variables):

$$L \equiv -\ln L^* = -n \ln(\alpha\beta) - (\beta - 1)\Sigma_i \ln t_i + \alpha T^\beta + \text{constant} \quad \text{(B-4)}$$

Since, in the Likelihood methods, monotonic transformations of the variables or Likelihood don't change the basic results, all the partial derivatives are with respect to $\ln \alpha$ and $\ln \beta$, because the results are more convenient. It is again straightforward to show that:

$$\frac{\partial L}{\partial \ln \alpha} = -n + \alpha T^\beta \quad \text{(B-4a)}$$

$$\frac{\partial L}{\partial \ln \beta} = -n - \beta\Sigma_i \ln t_i + \beta(\ln T)\alpha T^\beta \quad \text{(B-4b)}$$

The Maximum Likelihood solution is obtained by setting **Eq.** B-4 to zero and solving for $\hat{\alpha}$, $\hat{\beta}$; where ˆ denotes the Maximum Likelihood solution.

$$\hat{\beta} = n/\Sigma_i \ln(T/t_i) \quad \text{(B-5a)}$$

$$\hat{\alpha} = n/T^{\hat{\beta}} \quad \text{(B-5b)}$$

The uncertainties are evaluated by the Fischer Information matrix. For this purpose, the s-expected values of the second derivatives of L are needed.

Additional notation follows:

C = $\beta \ln T$

c = $\beta \ln t$

D = $EL_{11}EL_{22} - EL_{12}^2$

EL_{ij} = s-expected value of the second derivative of L with respect to the variables i and j; #1 corresponds to a, #2 corresponds to β

The second partial derivatives of L are:

$$L_{11} \equiv \frac{\partial^2 L}{\partial(\ln \alpha)^2} = \alpha T^\beta = m(T) \quad \text{(B-6a)}$$

$$L_{12} \equiv \frac{\partial^2 L}{\partial \ln \alpha \; \partial \ln \beta} = (\beta \ln T)\alpha T^\beta = Cm(T) \quad \text{(B-6b)}$$

$$L_{22} \equiv \frac{\partial^2 L}{\partial(\ln \beta)^2} = -\beta\Sigma_i \ln t_i + (\beta \ln T)\alpha T^\beta + (\beta \ln T)^2\alpha T^\beta$$

$$= -\Sigma_i \ln\left[\left(\frac{t_i}{T}\right)^\beta\right] + C[m(T) - n + Cm(T)] \quad \text{(B-6c)}$$

The random variables t_i and n appear only in Eq. B-6c. The s-expected value of n is $m(T)$ for a homogeneous or nonhomogeneous Poisson process. The s-expected value of $-\ln(t_i/T)^\beta$

$$-\int_0^T \ln\left(\frac{t_i}{T}\right)^\beta pdf\{t_i \mid n; T\} dt_i$$

$$= -\int_0^1 \ln x dx = +1 \tag{B-7}$$

where $x \equiv (t_i/T)^\beta$. Use EL_{ij} to denote the s-expected value of L_{ij} and take the s-expected values of Eqs B-6.

$$EL_{11} = m(T) \tag{B-8a}$$

$$EL_{12} = Cm(T) \tag{B-8b}$$

$$EL_{22} = m(T)(1 + C^2) \tag{B-8c}$$

$$D = [m(T)]^2 \tag{B-9}$$

Eq. B-8 contains the elements of the Fischer Information Matrix. The covariance matrix is the inverse of the Fischer Information matrix. Therefore the asymptotic covariance matrix for $\ln \alpha$ and $\ln \beta$ is

$$\text{AVar}(\ln \hat{\alpha}) = [1/m(T)](1 + C^2) \tag{B-10a}$$

$$\text{ACov}(\ln \hat{\alpha}, \ln \hat{\beta}) = -[1/m(T)]C \tag{B-10b}$$

$$\text{AVar}(\ln \hat{\beta}) = [1/m(T)] \tag{B-10c}$$

where AVar (•) and ACov (•) stand for the asymptotic variance and covariance, respectively.

In practice these are estimated by using the Maximum Likelihood estimates for α, β, and $m(T)$; namely, $\hat{\alpha}$, $\hat{\beta}$, n. ($\hat{C} \equiv \hat{\beta} \ln T$, $\hat{c} \equiv \hat{\beta} \ln t$).

The estimates for $m(t)$ and $\lambda(t)$ are

$$\hat{m}(t) = \hat{\alpha} t^{\hat{\beta}} = n(t/T)^{\hat{\beta}} \tag{B-11a}$$

$$\hat{\lambda}(t) = \hat{\alpha}\hat{\beta} t^{\hat{\beta}-1} = (\hat{\beta} n/T)(t/T)^{\hat{\beta}-1}$$
$$= \hat{\lambda}(T)(t/T)^{\hat{\beta}-1} \tag{B-11b}$$

$$\hat{\lambda}(T) \equiv \hat{\beta} n/T \tag{B-11c}$$

The uncertainty in $\hat{\lambda}$ is estimated by using the linearized expansion of $\ln\hat{\lambda}$, differentials, and the usual formula for variance of a linear function.

$$\ln \hat{\lambda}(t) = \ln \hat{\alpha} + \ln \hat{\beta} + (\hat{\beta} - 1)\ln t \tag{B-12a}$$

$$d \ln \hat{\lambda}(t) = d \ln \hat{\alpha} + d \ln \hat{\beta}$$
$$+ (\beta \ln t) d \ln \hat{\beta}$$
$$= d \ln \hat{\alpha} + (1 + c) d \ln \hat{\beta} \tag{B-12b}$$

$$[\hat{u}(t)]^2 \equiv \text{AVar}\{\ln \hat{\lambda}(t)\}$$
$$\approx \text{AVar}\{\ln \hat{\alpha}\} + (1 + c)^2 \text{AVar}\{\ln \hat{\beta}\}$$
$$+ 2(1 + c)\text{ACov}\{\ln \hat{\alpha}, \ln \hat{\beta}\}$$
$$= \{1 + [1 + \beta \ln (t/T)]^2\}/n \tag{B-13a}$$

$$\hat{u}(T) = \sqrt{2/n} \tag{B-13b}$$

where $\hat{u}(t)$ is an estimated, asymptotic standard deviation for $n(t)$.

The "point estimate plus or minus 2 standard deviations" is used to estimate the uncertainty in $\ln \hat{\lambda}(t)$. The factor of 2 is arbitrary; conventional wisdom in the USA often uses 3 standard deviations, but that seems excessive here. There is no "right" answer, nor does the uncertainty have an exact probabilistic interpretation. But it is a very useful tool for engineers and managers. Roughly speaking, the interval of uncertainty is such that no one has the vaguest idea where in the interval the true value lies, but you are quite sure it lies in, or reasonably close, to that interval.

The biggest reason that this concept is satisfactory is that often the region of uncertainty is so large that the main message to the user is that he knows much less than he hoped; or in a few cases, the region will be narrow enough that the user is quite satisifed, and would be even if the uncertainty were somewhat larger.

The upper and lower limits of the region of uncertainty for $\ell n\, \hat{\lambda}(t)$ are

$$\ln \hat{\lambda}_L(t) \equiv \ln \hat{\lambda}(t) - 2\hat{u}(t) \qquad \text{(B-14a)}$$

$$\ln \hat{\lambda}_U(t) \equiv \ln \hat{\lambda}(t) + 2\hat{u}(t) \qquad \text{(B-14b)}$$

or

$$\hat{\lambda}_L(t) = \hat{\lambda}(t) \exp\left[-2\hat{u}(t)\right] \qquad \text{(B-15a)}$$

$$\hat{\lambda}_U(t) = \hat{\lambda}(t) \exp\left[+2\hat{u}(t)\right] \qquad \text{(B-15b)}$$

The estimates are mathematically well behaved; e.g., it is easy to see that &, $\hat{\beta}$, $\hat{\lambda}_L(t)$, $\hat{\lambda}_U(t)$ are always positive.

This derivation has been for the case where the stopping rule is to stop after a fixed length of time (cumulative test time of all units). Maximum Likelihood results are often the same whether the stopping rule is for a "fixed time" or a "fixed number of failures". These estimates can be used regardless of which kind of stopping rule is invoked.

REFERENCE

1. E. Parzen, Stochastic Processes, Holden-Day, Inc. San Francisco, 1965.

INDEX

FOR THE COMMANDER:

ROBERT L. KIRWAN
Brigadier General, USA
Chief of Staff

OFFICIAL:

G. J. HAROLD
LTC, GS
Adjutant General

DISTRIBUTION:
Special

☆U.S. GOVERNMENT PRINTING OFFICE: 1976—603-770/309